

1 / 154

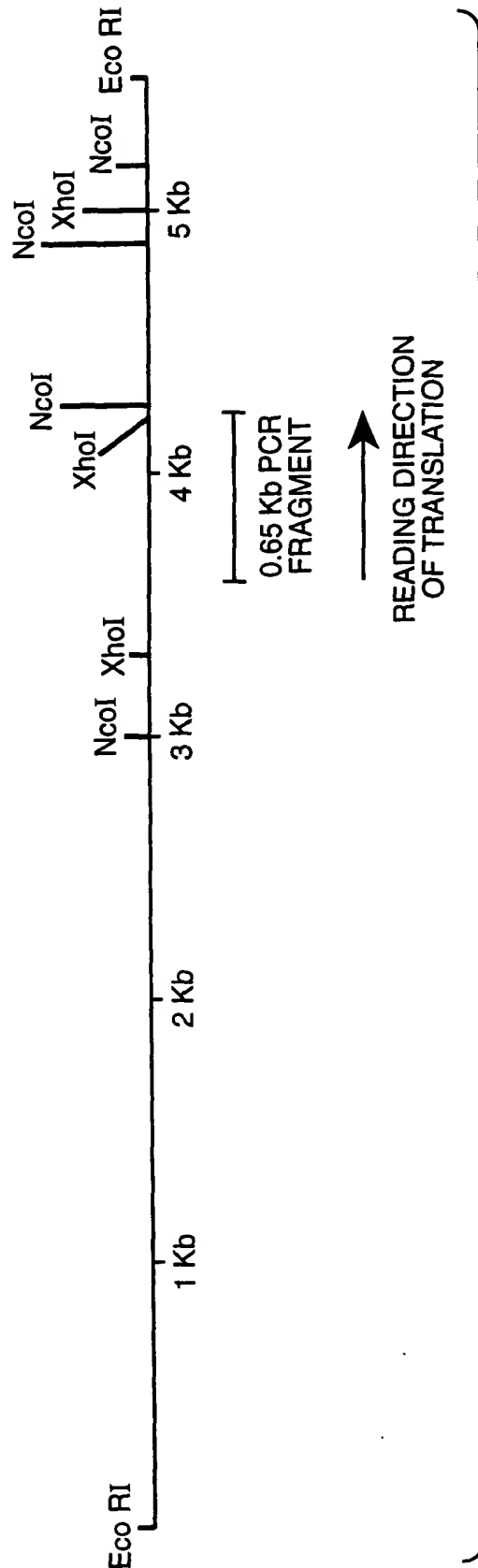
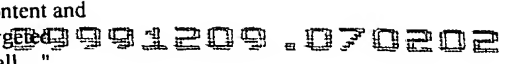


FIG. 1



content and
get 0.0001209 .070202
all "

3 / 154

Pst I Bpu10 I
ACGTCCTCGCAGTTGTGGTGAAGGACGCGCTTAGCAGCCTCTACGCAAGGCATCTCCGAAGACCTCTACAGCCGTTTAGTCGAAA 630

signal sequence

H V L A V V T A G H A L A A S T O G I S E D L Y S R L V E

Msc I Sal I
TGGCCACTATCTCCCAAGCTGCTACGCCGACCTGTGCAACATTCGGTCGACTATTATCAAGGGAGAGAAAAATTTACAATTTCTCAAACTG 720

M A T I S Q A A Y A D L C N I P S T I I K G E K I Y N S O T

BamH I BsaB I
ACATTACGGATGGATCCTCCGCGACGACGACGAGCAAGAAATAATCACCGTCTTCCGTCGCACTGGTAGTGATACGAATCTACAACCTCG 810

D I N G W I L R D D S S K E I I T V F R G T G S D T N L Q L

Eco31 I
ATACTAACTACACCCCTCAGCCTTTCGACACCCCTACCACAATGCAACGGTTGTGAAGTACACGGTGGATATTATATTGGATGGGTCCTCCG 900

D T N Y T L T P F D T L P Q C N G C E V H G G Y Y I G W V S

FIG._2B

4 / 154

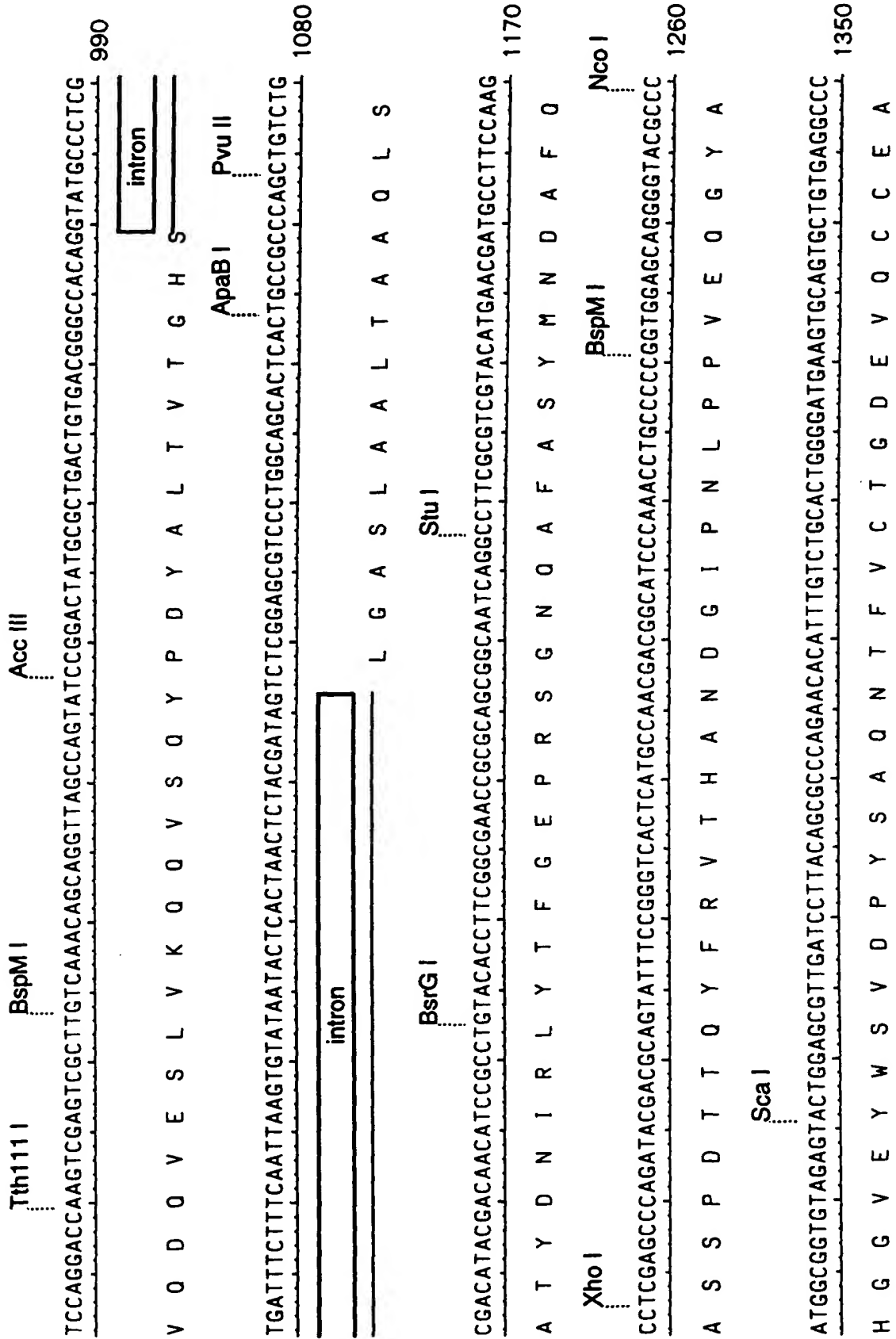


FIG. 2C



6 / 154

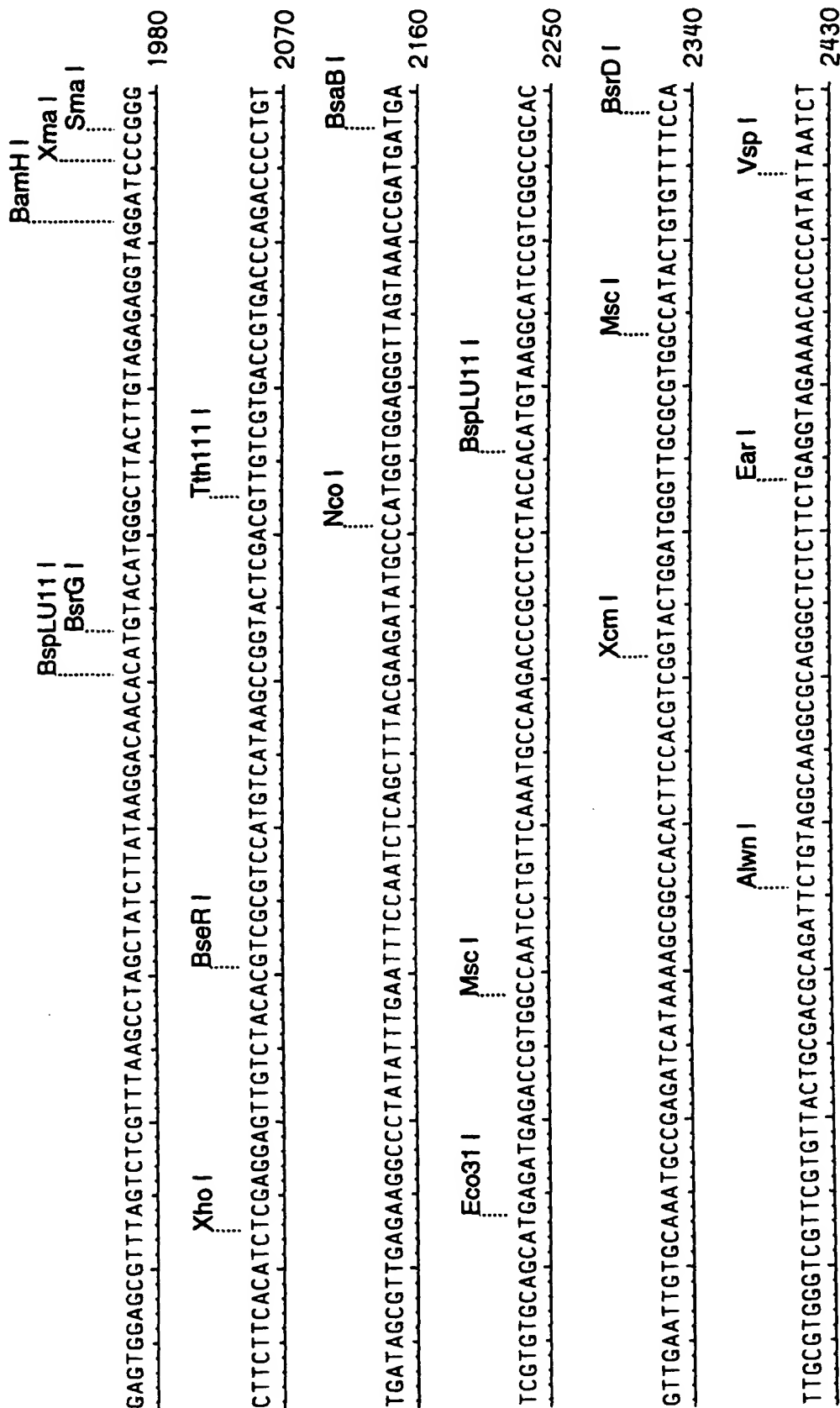


FIG.-2E



7 / 154

CCATGGTGTGCGATATCGGCAGTAGTCTTTGCGGAAACGTTGAGGGTTACAGTGATCTGCGTGGACATACATT
CGGGGAATCTACGGCGGAATATCAAGTCTTTCGGAATATCCATATTTGGAAAGGACAGAAAGCTCCGGGGTAGTTT
GATAGTAGCTCCGGTGATTAATTCGGGAGCTGACAGGAGTGAGCGTCACTGTAGACCATCTAGTAATGTTCAGT
CGCGCGCAATTCGCACATGAACAAGTTGATTTTCGGGACCCCATTTGTTACATCTCTCGGCTACAGCTCGAGATG
TGCTGCGGAGTATCTTAGAGAGCCATGCCAGCGTGTGTTTATACGACCAAAAGTCAAGGAAATATGAACGATCG
TCGGATATTTCTTGTGTTTATCCTAAATAGTCTTCAGTGGTTTATTAAGAGATAGATCCCTTCACAAACACT
CATCCAACGGACTTCTCATACCACTCATTTGACATAAATTTCAAAACAGCTCCAGGCGCATTTAGTTCAACATGAAGC
AATTCCTCCGCCAAACACGTCCTCGCAGTTGTGGTGACTGCAAGGCAAGCTTACGAGCTCTACGCAAGGCATCT
CCGAAGACCTCTACAGCCGTTTAGTCGAAATGGCCACTATCTCCCAAGCTGCCACGCCGACCTGTGCAACATTC
CGTCGACTATTAACAAGGAGAGAAATTTACAATCTCAAACTGACATTAACGGATGGATCTCTCCGACGACA
GCAGCAAGAATAAATCACCGTCTTCCGTCGCACTGGTAGTACGAACTCAAACTCGATACTAACATACACCC
TCACGCCCTTCGACACCCCTACCAATGCAACGGTTGTGAAGTACACGGTGGATATATATTGGATGGGTCTCCG
TCCAGGACCAAGTCGAGTCGCTTGTCAAAACAGCAGGTTAGCCAGTATCCGGACTATGCGCTGACTGTGACGGGCC
ACAGGTATGCCCTCGTGATTTCTTCAATTAAGTGTAATAACTCACTAACTCACTACGATAGTCTCGGAGCGTCCC
TGGCAGCACTCACTGCCGCCAGCTGTCTGCGACATACGACAACTCCGCCCTGTACACCTTCGGCGAAACCGCGCA
GCGGCAATCAGGCCCTTCGCGTCGTACATGAACGATGCCCTTCAAGCCCTCGAGCCAGATACGACGCAATATTTCC
GGGTCACTCATGCCAAACGACGGCATCCCAAACTGCCCTGGAGCAGGGGTACGCCCATGGCGGTGTAGAGT
ACTGGAGCGTTGATCCTTACAGCGGCCCAAGAACACATTTGTCTGCACTGGGGATGAAGTGCAGTGTGTGAGGCC
AGGCGGACAGGGTGTAATAATGCGCACACGACTTATTTTGGGATGACGAGCGGAGCTGTACATGTTGATCAG
TCATTTACGCCCTCCCGAGTGATCCAGGAAGATGGATGTCTGGAGAGGGCATGTCATGTACGTATACCCGAAAGC
ACACTTTTTCGGTAATCAGGACATGTAATAAGTTCCCTTCATGAATAGATATGGTTACCTCACCATAAAGCCTT
GAGGTGCTTCTCTTTGATTTGTGAATATATATTTAAAGTAGATGACAGATATCTCTAAACACCTTATCCGCT
TAAACCCATCATAGATTGTGTACAGTGATAGACCCCTTGAATGATGAGCGAAATGTATCAGTCCCGTTTAAATCA
AACCCTTTCAGCTAGCACAGTCAGAAATACACCAACCCCATTTCTAAGGTAGTACTAAATATGAATACAGCCTAAA
TGCAATCGCTATATGATCCCATAAAGAGCAACAACTTTCAGATCTCGTTTTCGCTGCGAAGAGCTAGCTCTAC
CATGGTCTCAATTATGAGTGGAGCGTTTAGTCTCGTTTAAAGCCTAGCTATCTTATAAGGACAACACATGTACATG
GGCTACTGTAGAGGTAGGATCCCGGGCTTCTTCACATCTCGAGGAGTTGTCTACACGTCGCGTCCATGTCA
TAAGCCGGTACTCGACGTTGTCTGTGACCGTGACCCAGACCCCTGTGATAGCGTTGAGAGGCCCTATATTGAA
TTTCCAAATCTCAGCTTACGAAAGATATGCCCATGGTGAGGGTTAGTAAACCCGATGATGATCGTGTGCAGCATGA
GATGAGACCGTGGCCAAATCCCTGTTCAAAATGCCAAAGACCCGCCCTCCACCATGTAAAGCATCCGTCGGCCGCAC
GTTGAATTGTGCAATGCCGAGATCATAAAGCGGCCACACTTCCACGTCGGTACTGGATGGGTGCGCGTGGCC
ATACGTGTTTCCATGCGTGGGTGTTCTGCGACGCAATTTCTGTAGGCAAGGCGCAGGCGCTCTCT
TCTGAGGTAGAAACACCCCATATTAATCTGAATC

FIG._3

8 / 154

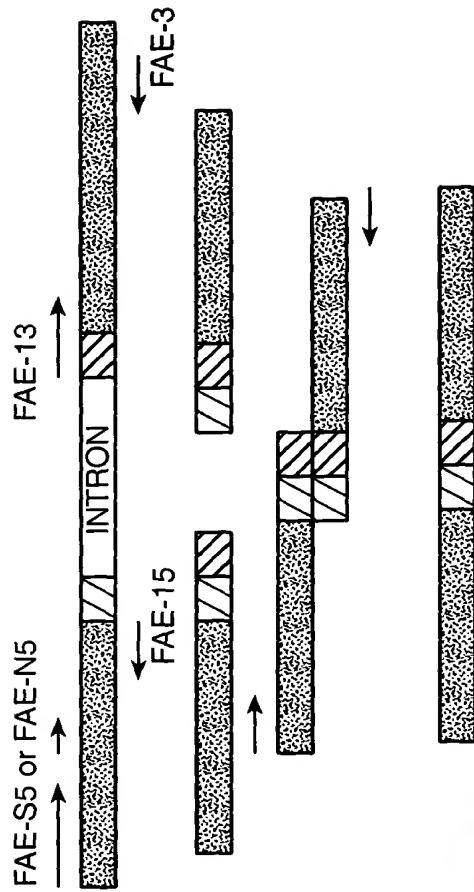


FIG. 4

FAE-I3 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCACTC 35-mer
 FAE-I5 GCGCCGAGGAGTGCGCGGTACCGTCAGCGCGTAGTCC 40-mer

intron position in original

complement, FAE-I5
 FAE-I3

Y A L T V T G H S L G A S L A A L
 GGACTACGCGCTGACCGTACCGGCCACTCCCTCGGCGCC
 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCACTC
 Y A L T V T G H A L G A S L A A L

FIG. 5

9 / 154

Vector Construction

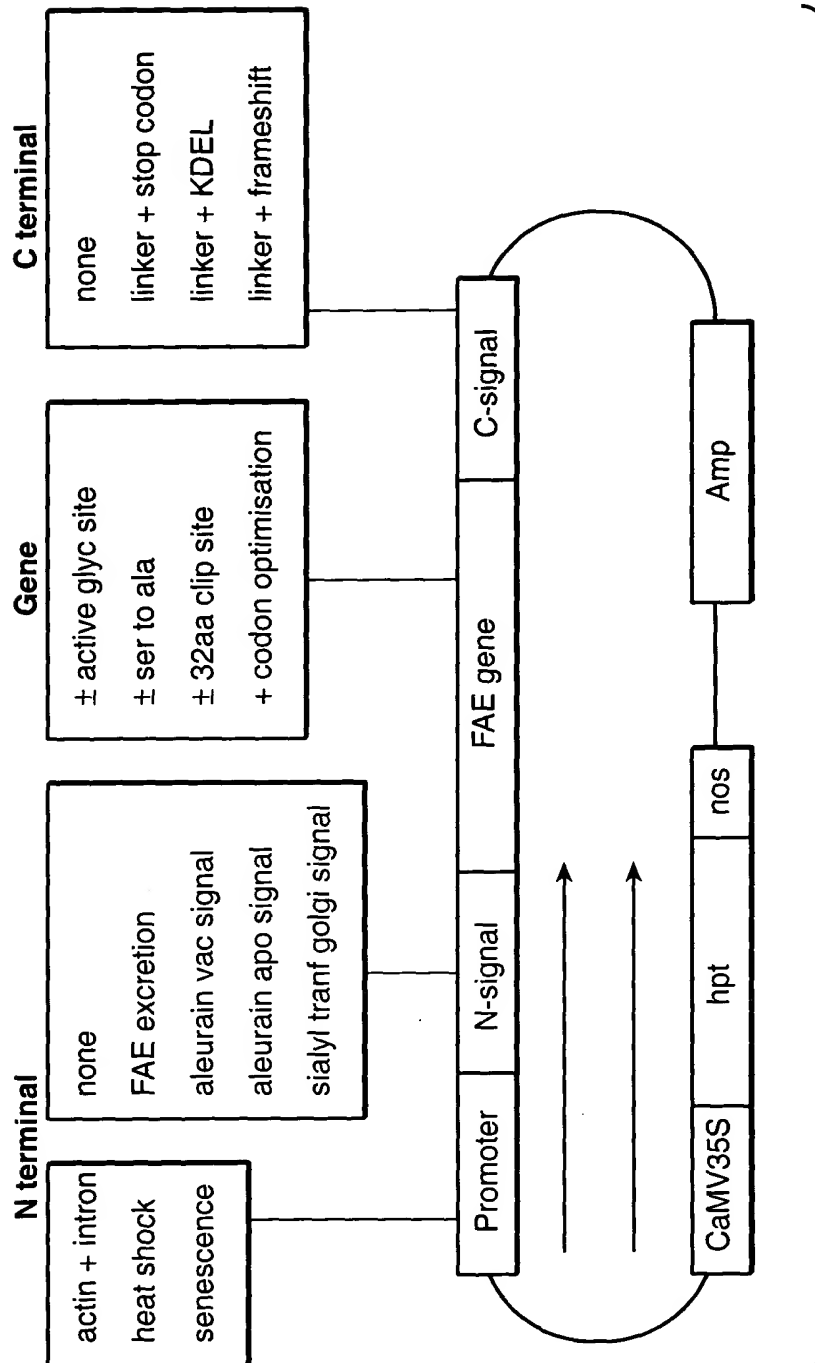


FIG._6

10 / 154

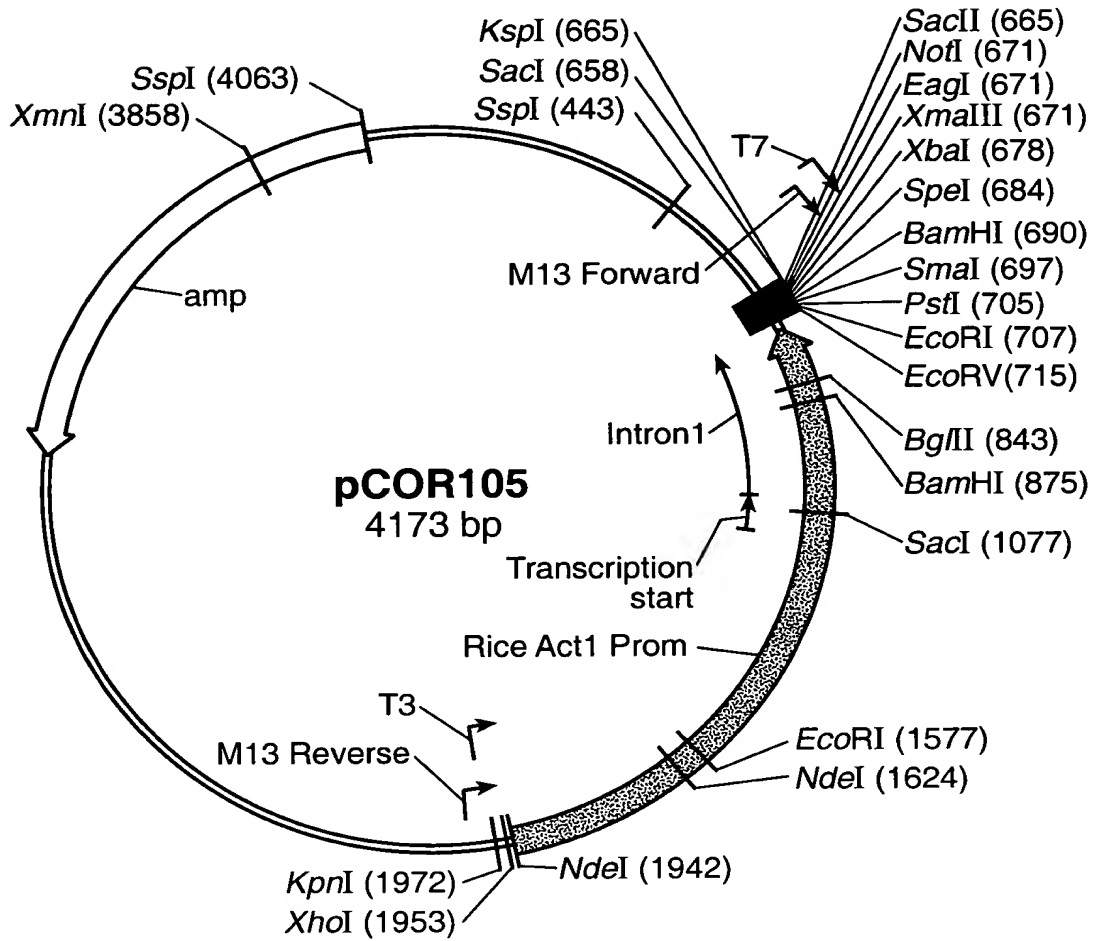


FIG._7

11 / 154

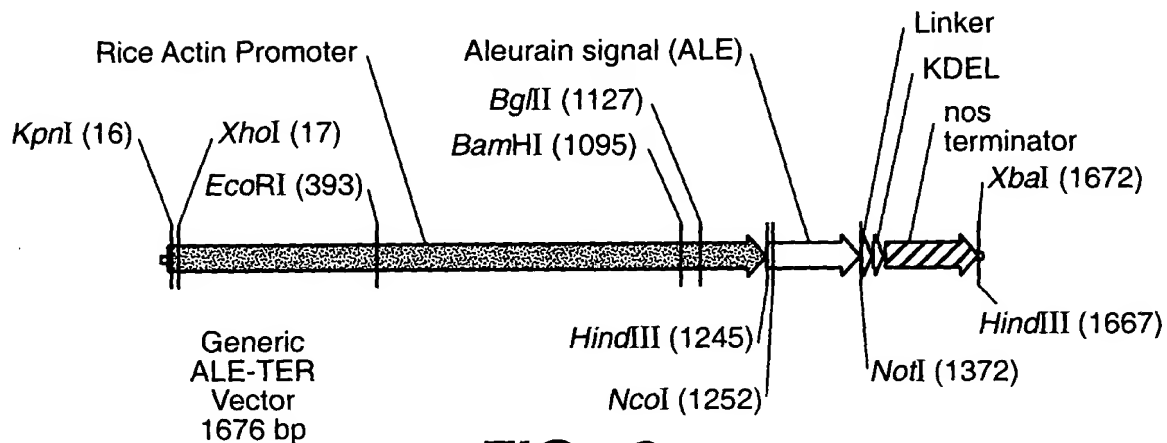


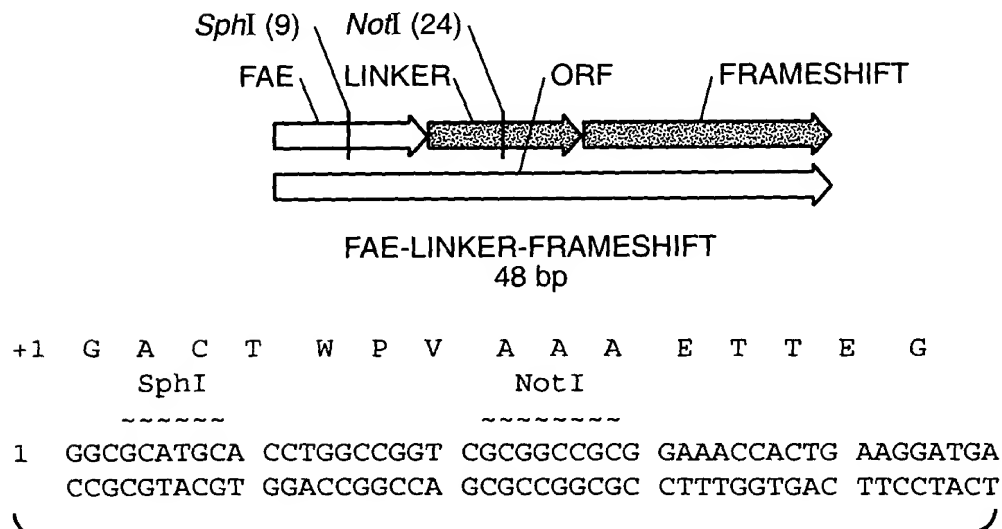
FIG._8

KDEL-COOH ER retention sequence

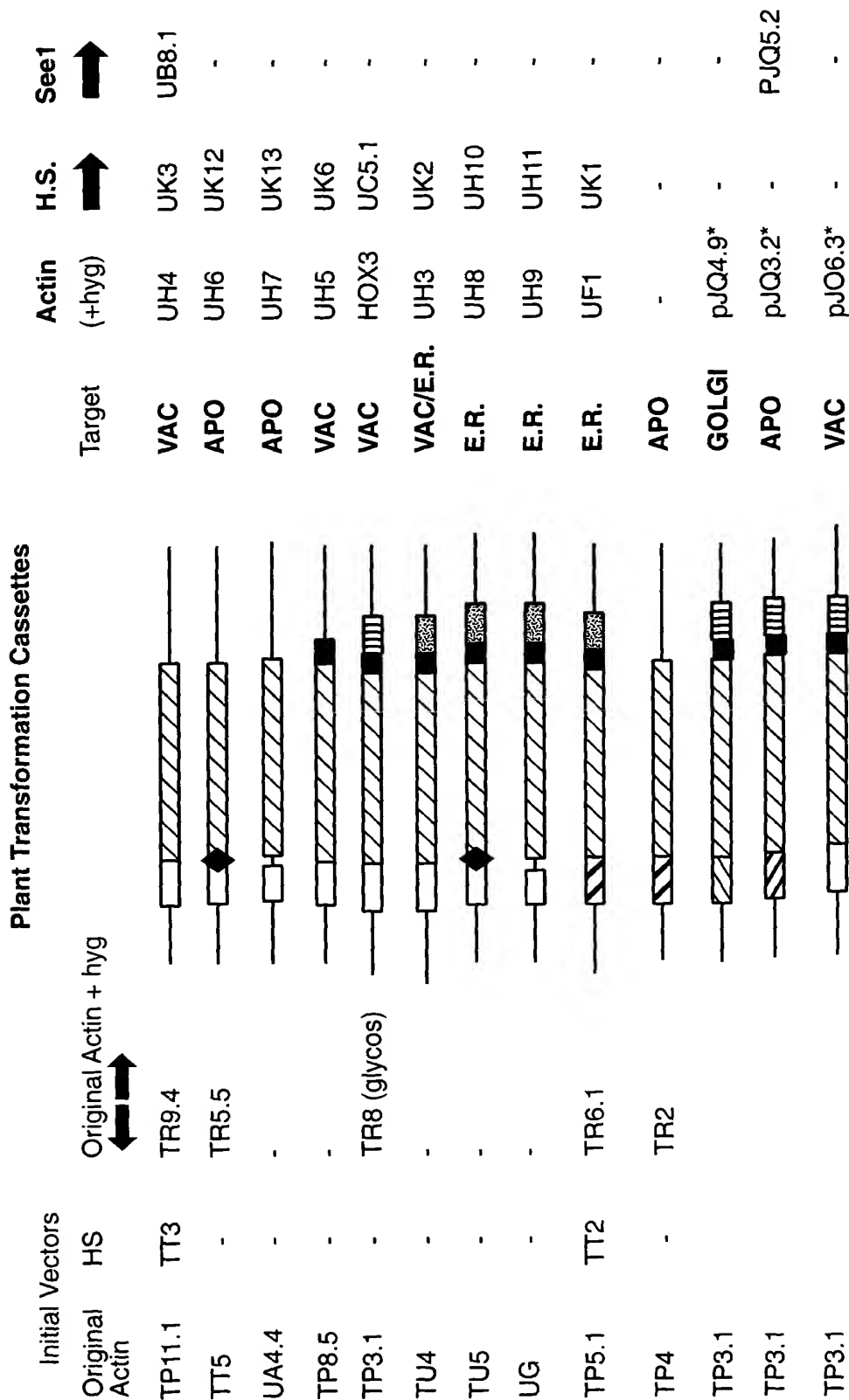
NotI
 ~~~~~  
 A A A K P L K D E L \*  
 1 GCGGCCGCGA AACCACTGAA GGATGAGCTG TAA

**FIG.\_9**

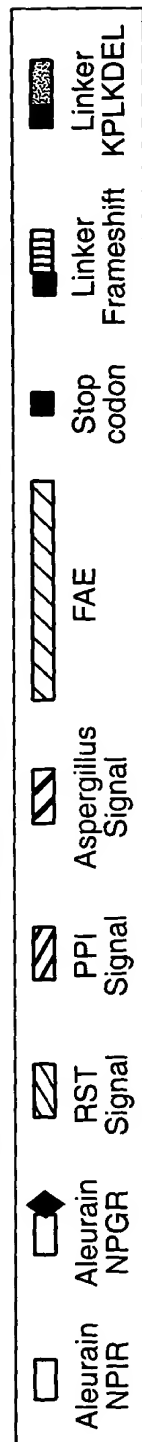
**FAE-LINKER-FRAMESHIFT Structure and Sequence**



**FIG.\_10**



\* - Modified Actin Promoter (Kpn1-EcoR1 Deletion and Restored NCO Site)



**FIG.-11**



13 / 154

## Vectors

### Original Actin promoter in pCOR105

|       | Target | Signal sequences                      | Vectors                      |
|-------|--------|---------------------------------------|------------------------------|
| (i)   | APO    | - aleurain-NPGR-FAE                   | pUH6, pTT5, TT5.5, pTT5.1    |
|       |        | - aleurain-delNPIR -FAE               | pUH7, pUA4.4,                |
| (ii)  | ER     | - aleurain-NPGR-FAE-linker-KDEL       | pTU5, pUH8,                  |
|       |        | - aleurain-delNPIR-FAE-linker-KDEL    | pUG4, pUH9,                  |
| (iii) | VAC    | - aleurain-NPIR-FAE                   | pTP11.1, pTR9.4, pUH4, pUK3, |
| (iv)  | ER/VAC | - aleurain-NPIR-FAE-linker-KDEL       | pTU4, pUH3,                  |
| (v)   | VAC    | - aleurain-NPIR-FAE-linker-frameshift | pUA1K3, pTP3.1, pUC5.11      |
| (vi)  | VAC    | - aleurain-NPIR-FAE-linker-stop       | pTP8.5, pUH5                 |
| (vii) | ER     | - Aspergillus signal -FAE-KDEL        | pTP5.1, pTP6.1, pUF1,        |

### Modified actin promoter (Kpn1-EcoR1 deletion and restored NCO site)

|       |       |                                       |        |
|-------|-------|---------------------------------------|--------|
| (i)   | VAC   | - aleurain-NPIR-FAE-linker-frameshift | pJ06.3 |
| (ii)  | GOLGI | - RST-FAE-linker-frameshift           | pJQ3.2 |
| (iii) | APO   | - PPI-FAE-linker-frameshift           | pJQ4.9 |

### Heat-shock promoter

|       |        |                                       |                  |
|-------|--------|---------------------------------------|------------------|
| (i)   | APO    | - aleurain-NPGR-FAE                   | pUH12            |
|       |        | - aleurain-delNPIR-FAE                | pUH13            |
|       |        | - Aspergillus signal-FAE              | pTP4a2, pTR2.22, |
| (ii)  | ER     | - aleurain-NPGR-FAE-linker-KDEL       | pUH10            |
|       |        | - aleurain-delNPIR-FAE-linker-KDEL    | pUH11            |
| (iii) | VAC    | - aleurain-NPIR -FAE                  | pUK3, pTT3       |
| (iv)  | ER/VAC | - aleurain-NPIR-FAE-linker-KDEL       | pUK2             |
| (v)   | VAC    | - aleurain-NPIR-FAE-linker-frameshift | pUC5.11, pHOX3   |
| (vi)  | VAC    | - aleurain-NPIR-FAE-linker-stop       | pUK6             |
| (vii) | ER     | - Aspergillus signal -FAE-KDEL        | pUK1, pTT2       |

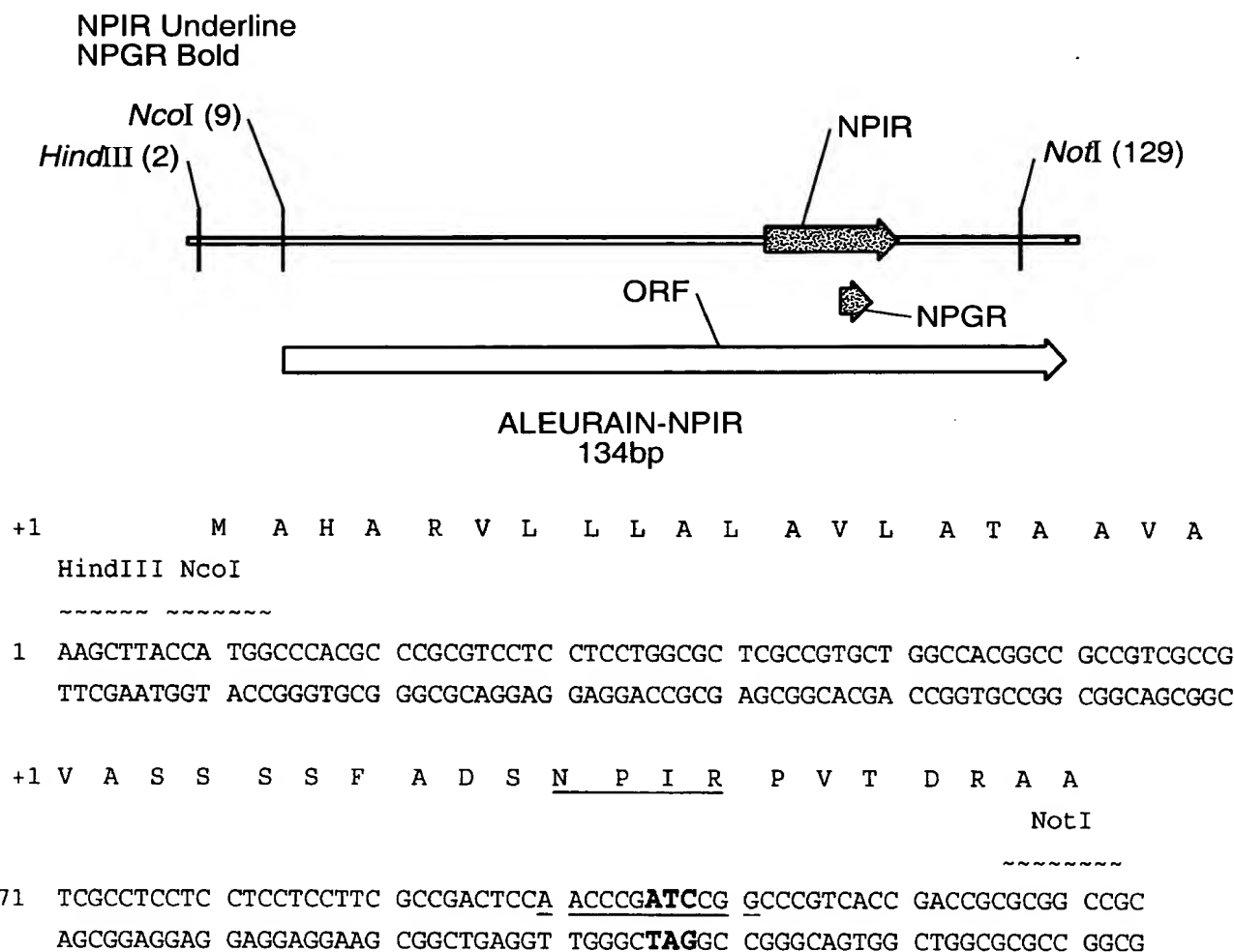
### Senescence promoter

|      |     |                                  |        |
|------|-----|----------------------------------|--------|
| (i)  | APO | - See1-PPI-FAE-linker-frameshift | pJQ5.2 |
| (ii) | VAC | - See1-aleurain-deleted NPIR-FAE | pUB8.1 |

**FIG.\_12**

14 / 154

# **ALEURAIN-NPIR (Vacuolar) and NPGR (Apoplast) Structure and Sequence**



**FIG. 13**

# RAT SIALYL TRANSFERASE Golgi signal sequence

```

HindIII
~~~~~
 M I H T N L K K K F S L F I L V F L L F A
1 AAGCTTACCA TGATCCACAC CAACCTCAAA AGAAGTTCT CCTCTTTCAT CCTCGTCTTC CTCCTCTTCG

 . V I C V W K K G S D Y E A L T L Q A K E F Q M
71 CCGTGATCTG CGTGTGGAAG AAGGCTCCG ACTACGAGGC CCTCACCCTC CAAGCCAAGG AGTCCAAAT

 NotI
      ~~~~~
      . A A
141  GGCGGCCCGC
  
```

FIG.\_14

# POTATO PROTEASE INHIBITOR II Apoplast signal sequence

```

HindIII
~~~~~
 M X V H K E V N F V A Y L L I V L G L L L
1 AAGCTTACMA TGMCGTGCA CAAGGAGGTS AACTTCGTSG CCTACCTCCT GATCGTSCTC
 GGCCCTCCTCT

 NcoI
      ~~~~~
      . L V S A M E H V D A K A C T X E C G N L
      G F G .
71  TGCTCGTSTC CGCCATGGAG CACGTGGACG CCAAGGCCTG CACCCKCGAG TCGGGCAACC
      TCGGCTTCGG

      NotI
      ~~~~~
 . I C P A A A
141 CATCTGCCCG GCGGCCGCC

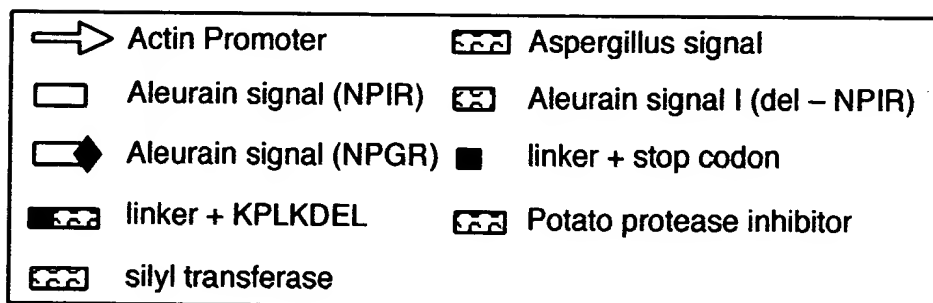
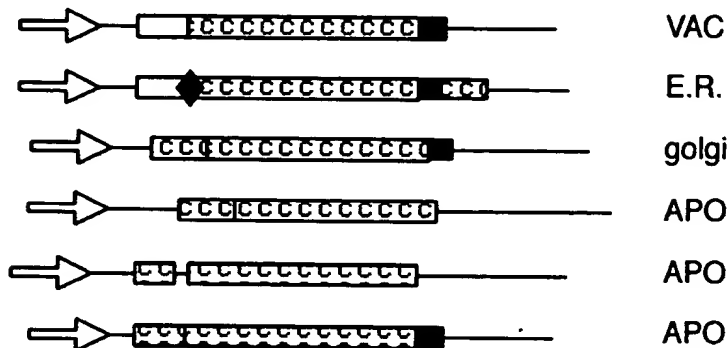
```

FIG.\_15

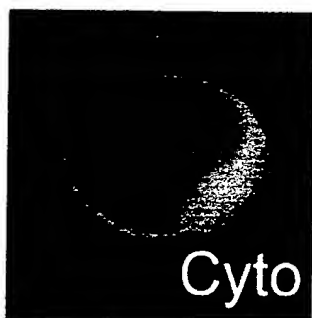
16 / 154

# Targeting Expression of gfp to Different Cell Compartments

## Actin Promoter Targeting Vectors



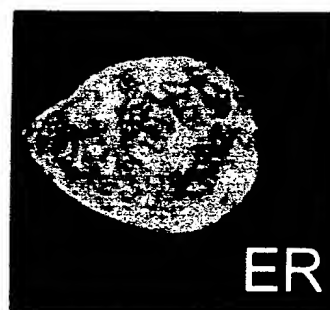
**FIG.\_16A**



**FIG.\_16B**



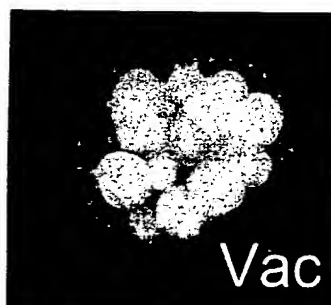
**FIG.\_16C**



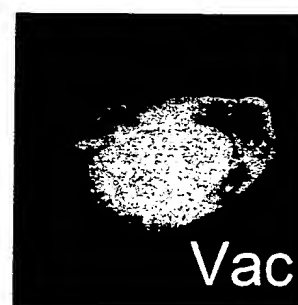
**FIG.\_16D**



**FIG.\_16E**



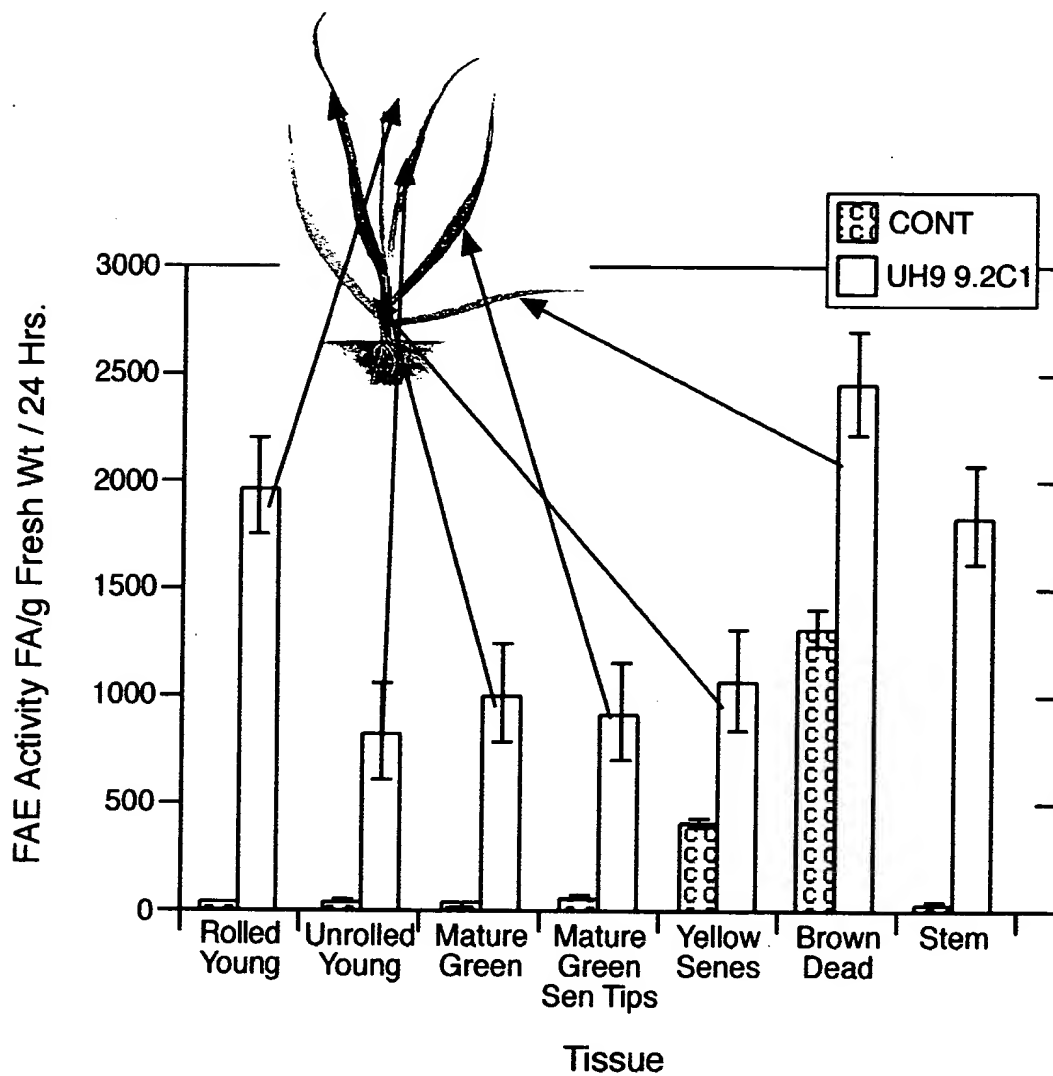
**FIG.\_16F**



**FIG.\_16G**

17 / 154

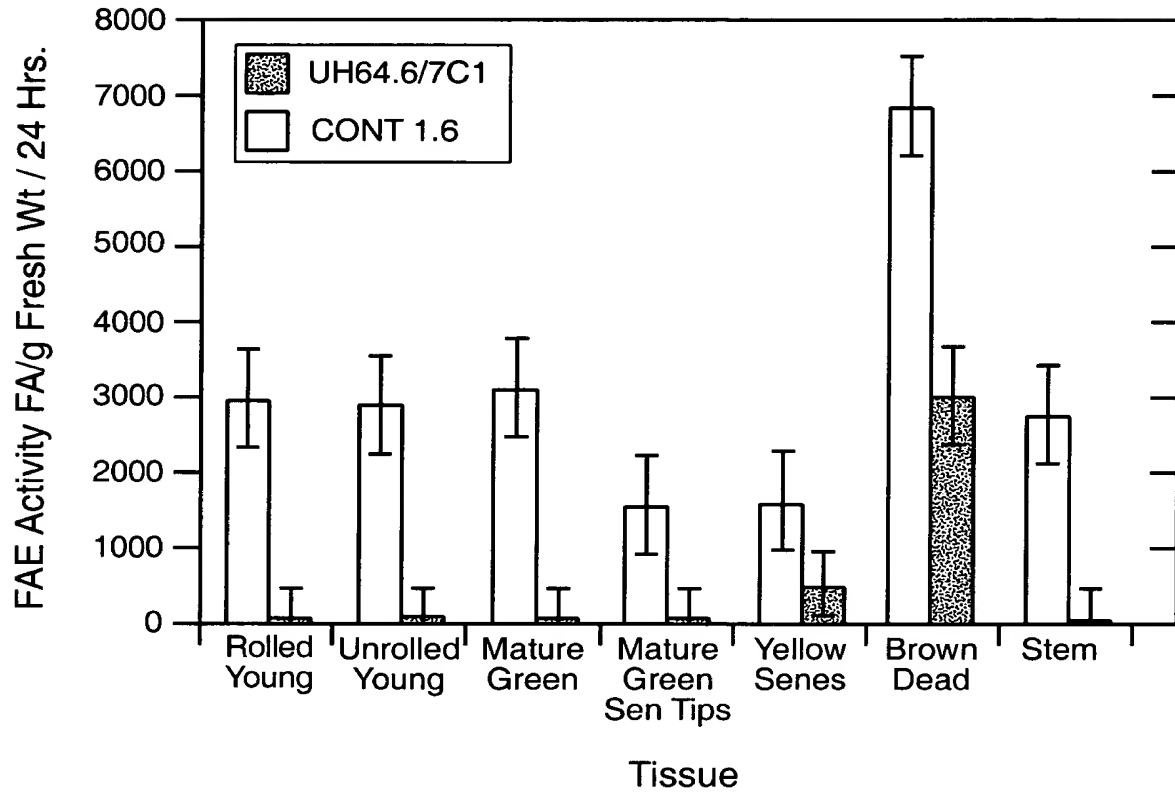
**FAE Activity in Transgenic *Festuca arundinacea* Leaves  
of Different Ages Under ER and APO Targeting Sequence**



**FIG. 17A**

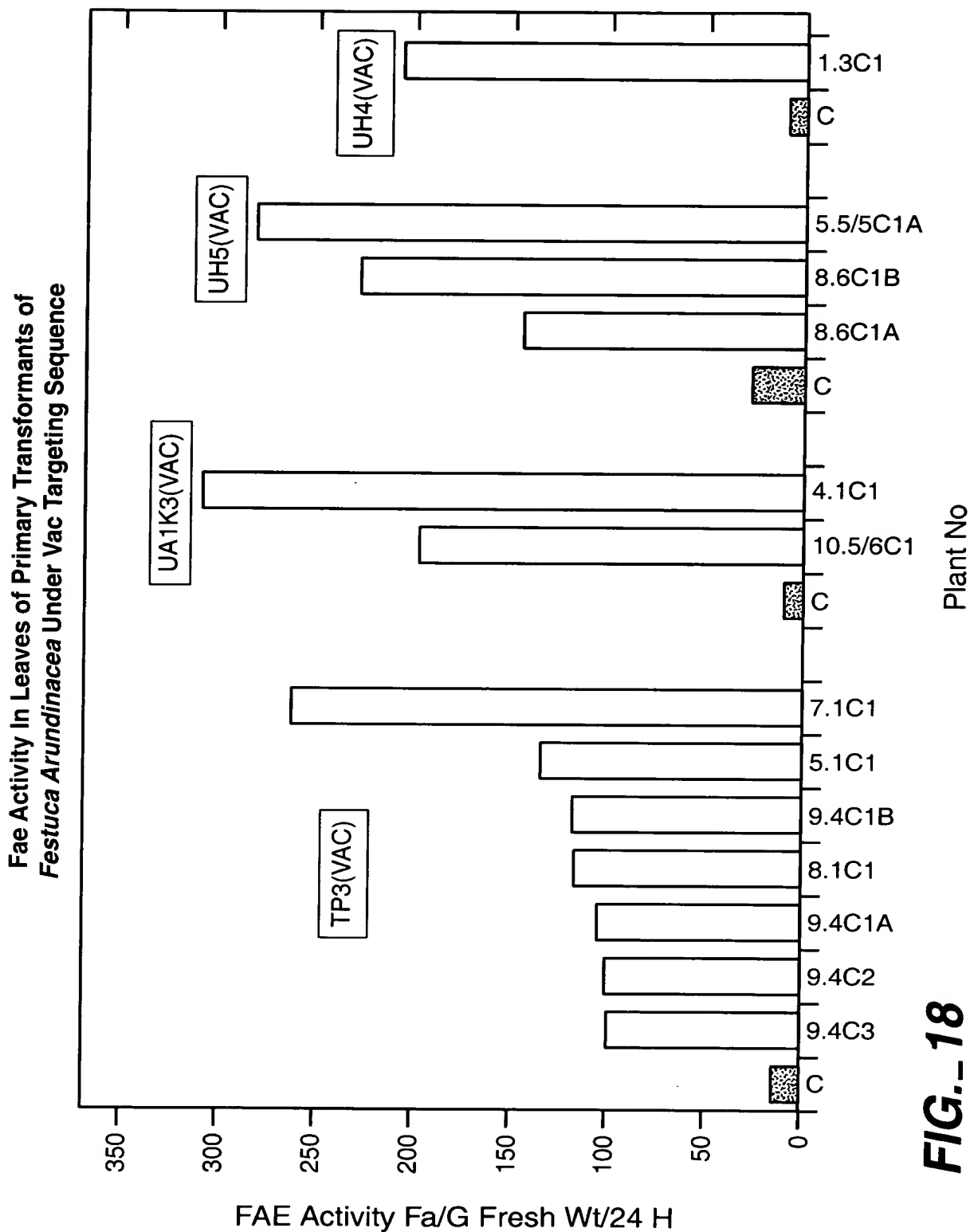
18 / 154

**FAE Activity in Transgenic *F. stuebelii* Leaves of Different Ages Under ER and APO Targeting Sequence**



**FIG. 17B**

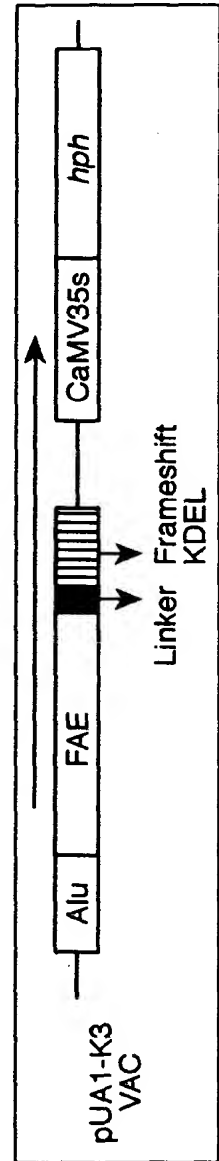
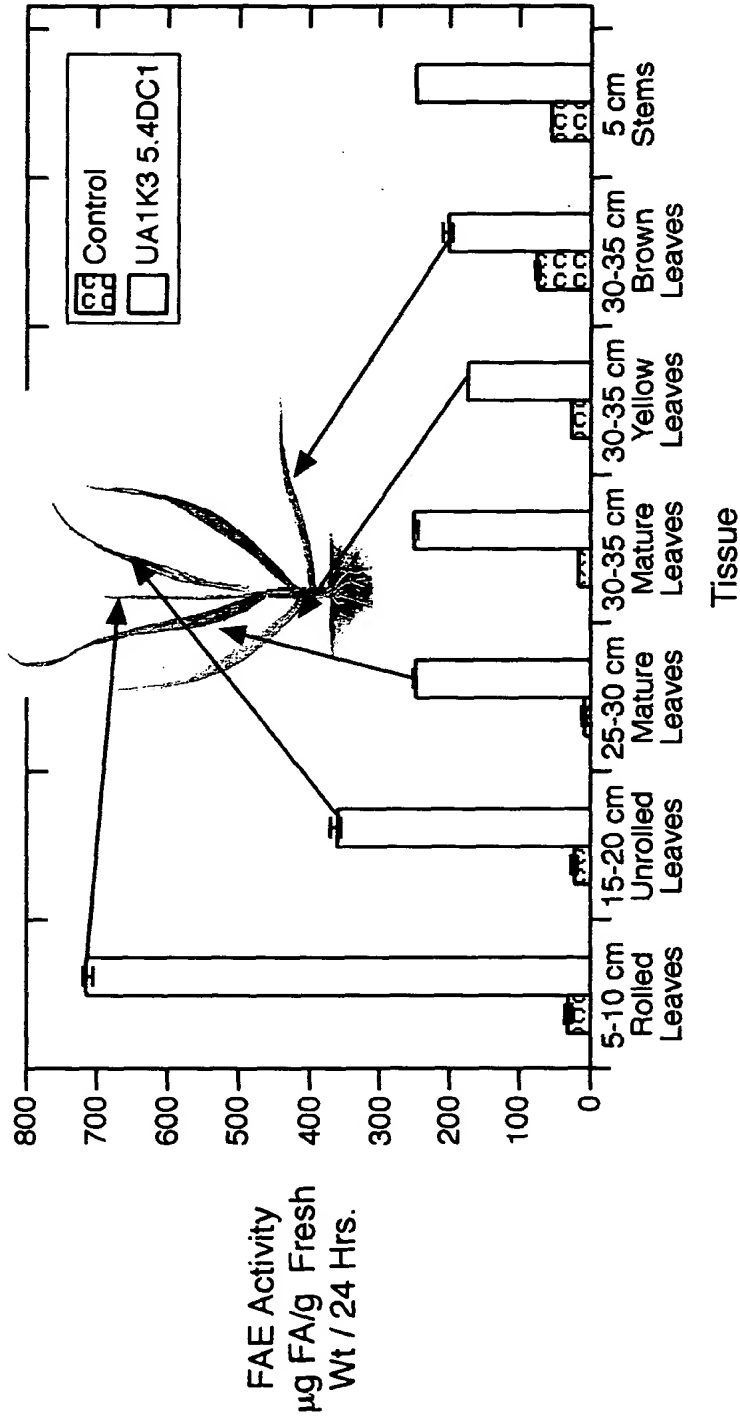
19 / 154



20 / 154

FIG. 19

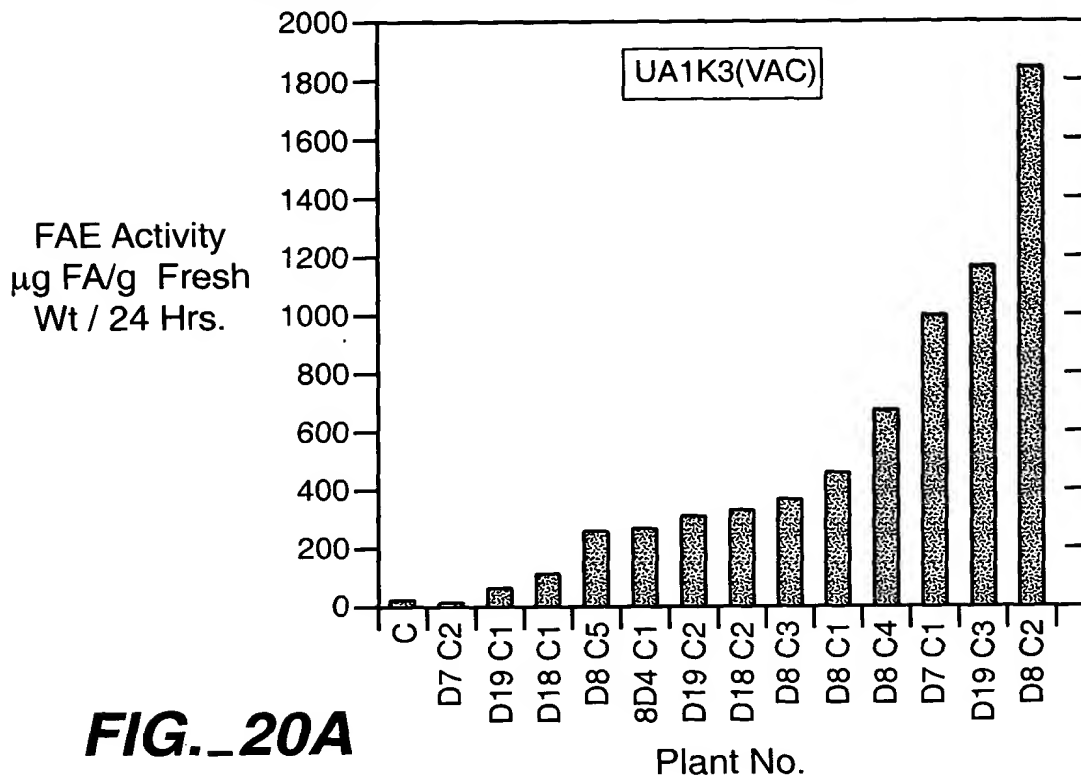
FAE Activity in *Lolium multiflorum* Leaves of Different Ages





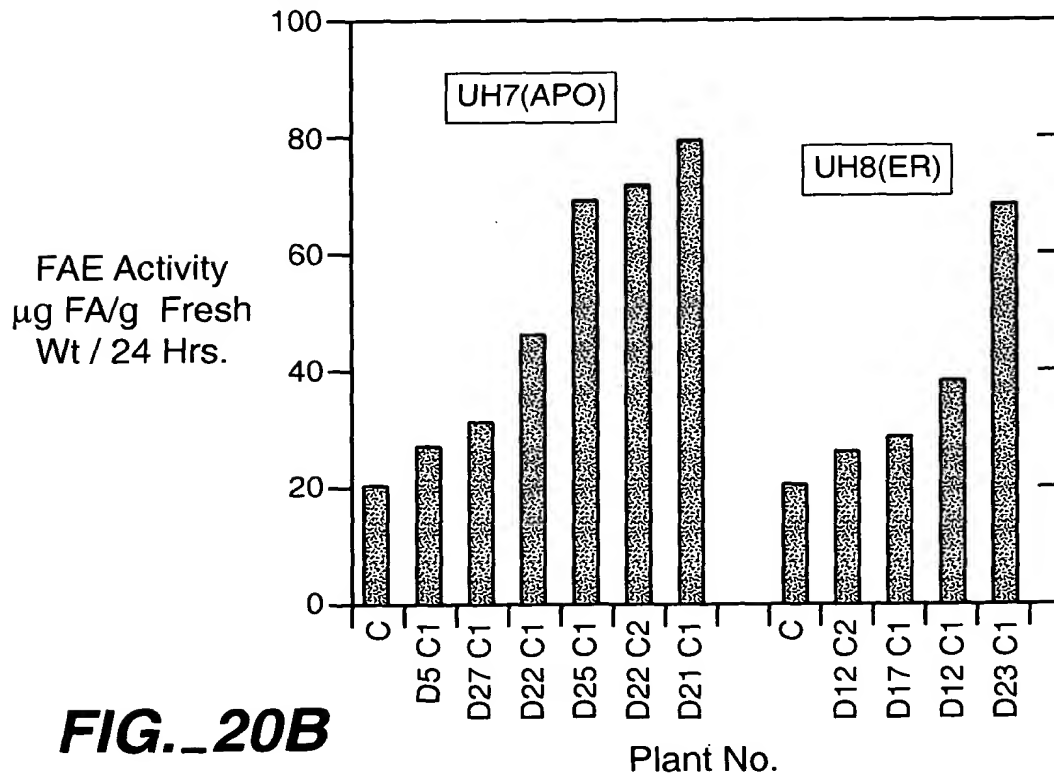
21 / 154

**FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence**



**FIG.\_20A**

**FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence**



**FIG.\_20B**

Levels of Esterified Monomeric and Dimeric Hydroxycinnamicacids in *Festuca Arundinacea* Plants Expressing FAE Under VAC Targeting Sequence

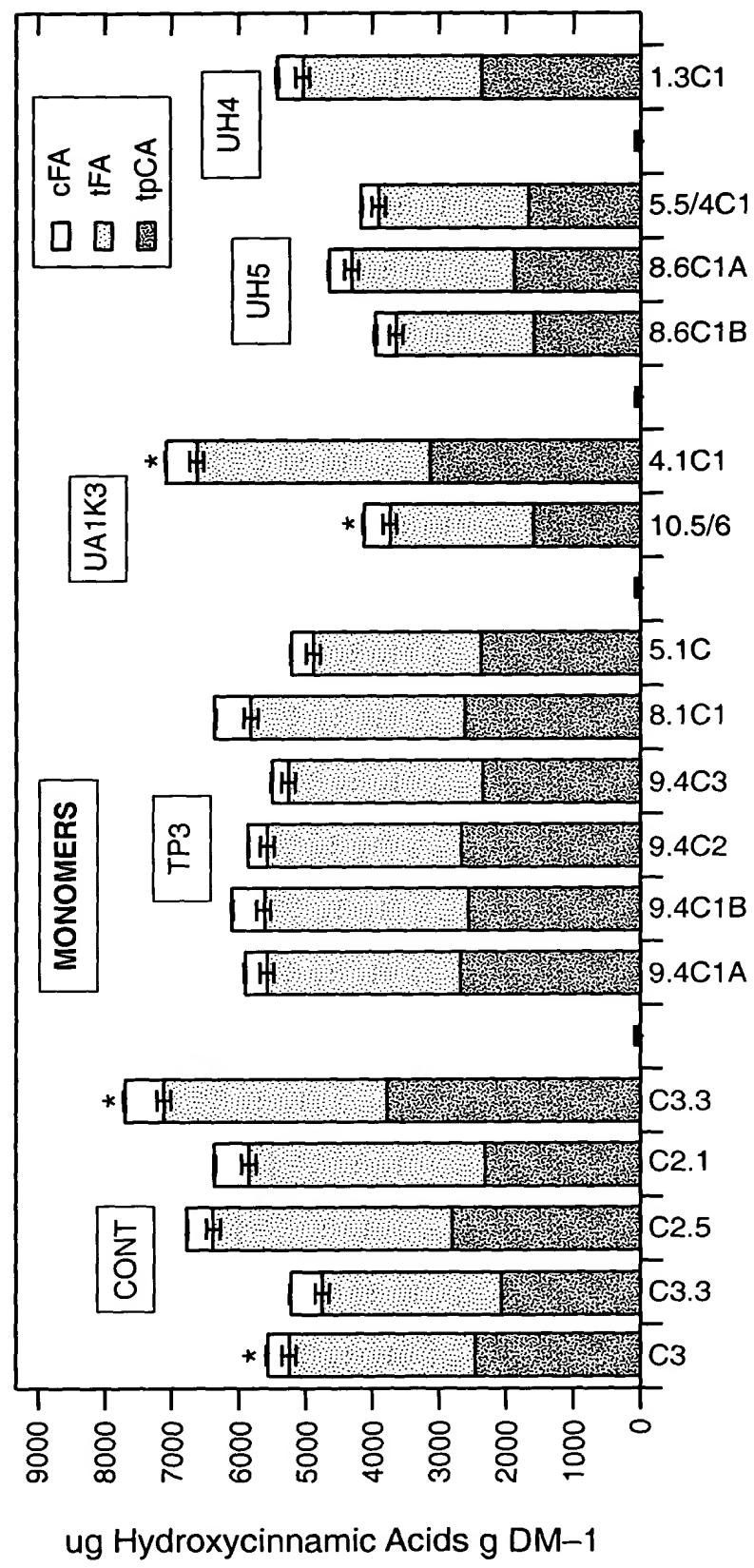


FIG..21A

23 / 154

Levels of Esterified Monomeric and Dimeric Hydroxycinnamicacids in  
*Festuca Arundinacea* Plants Expressing FAE Under VAC Targeting Sequence

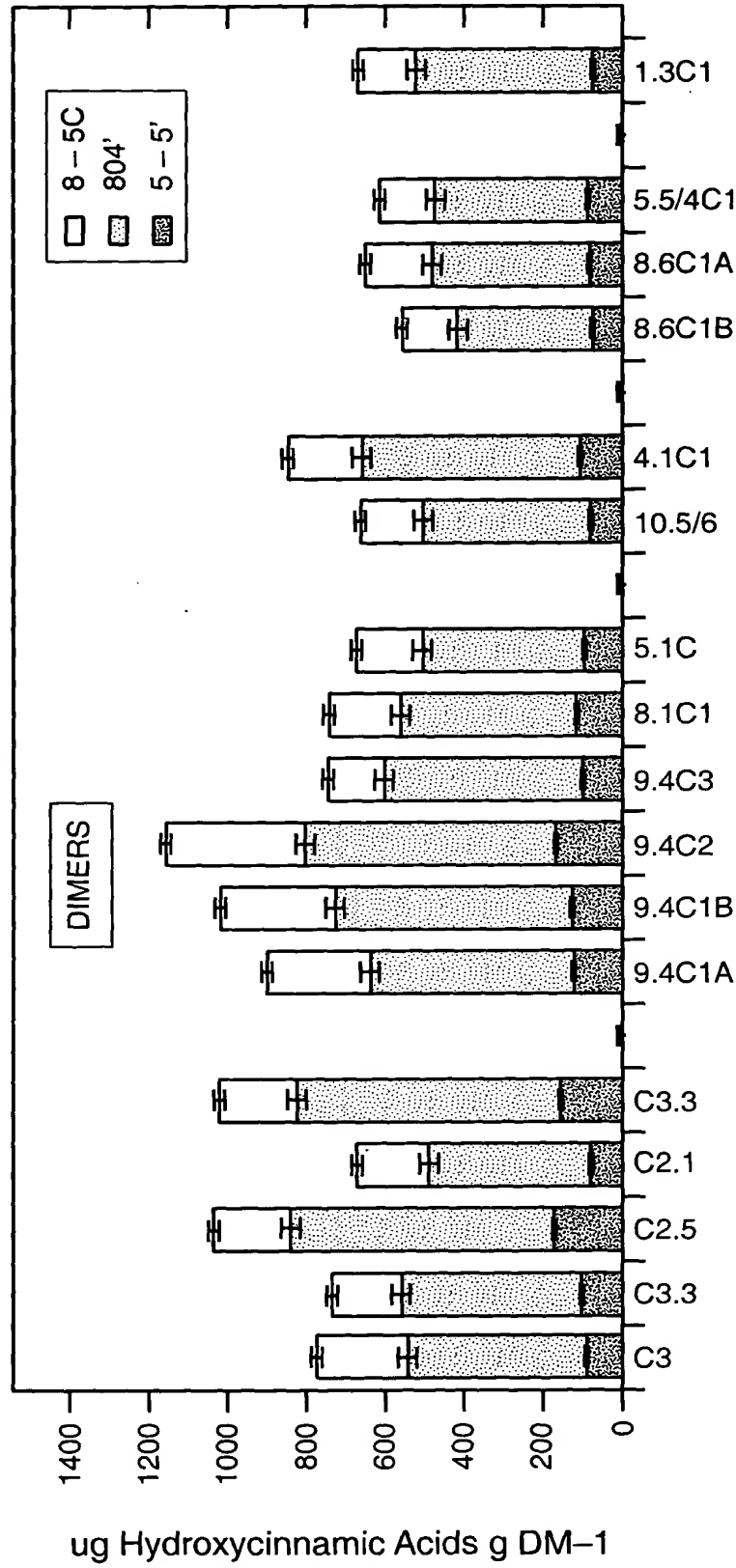


FIG. 21B

24 / 154

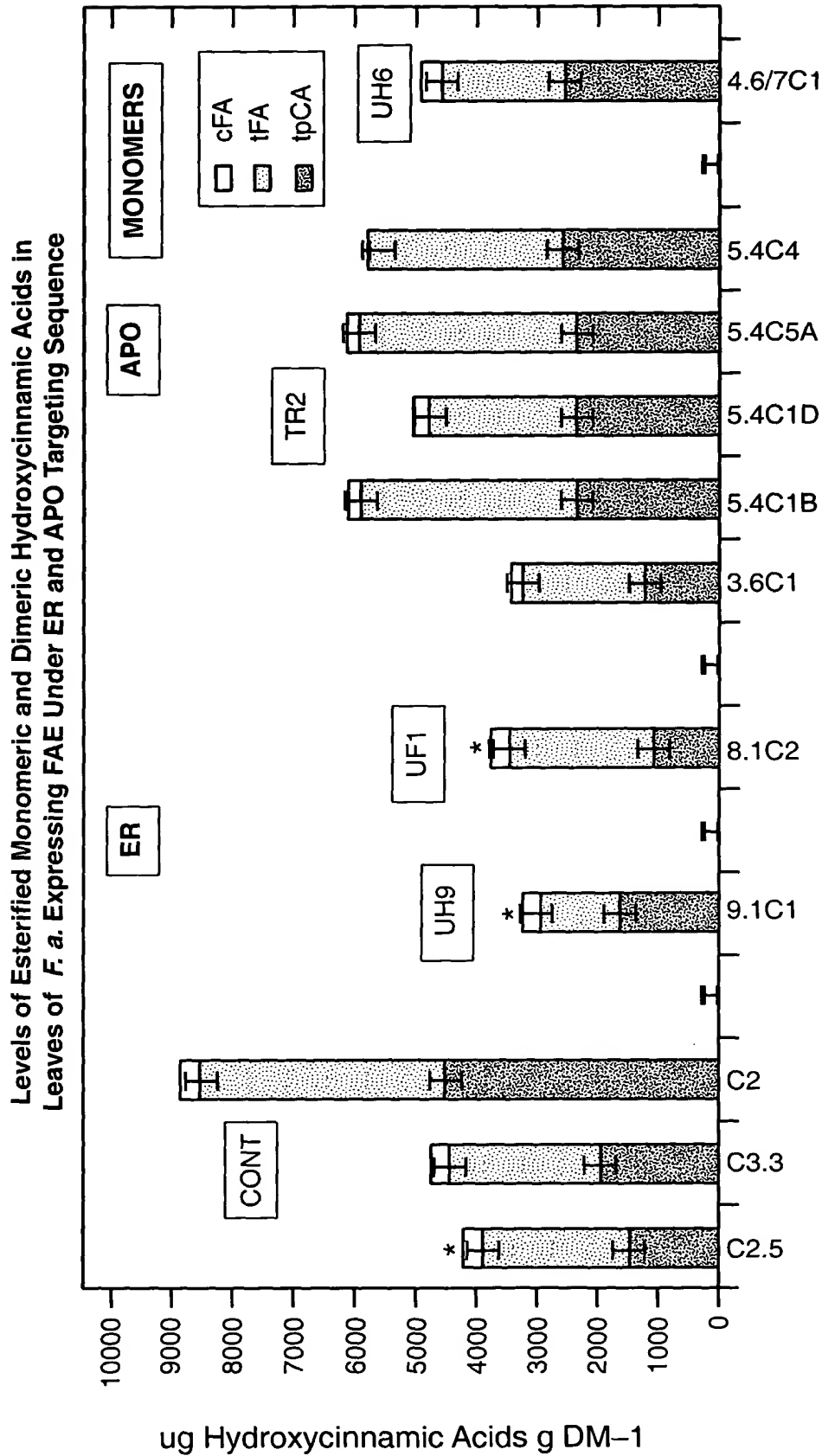


FIG.\_22A

25 / 154

Levels of Esterified Monomeric and Dimeric Hydroxycinnamic Acids in Leaves of *F. a.* Expressing FAE Under ER and APO Targeting Sequence

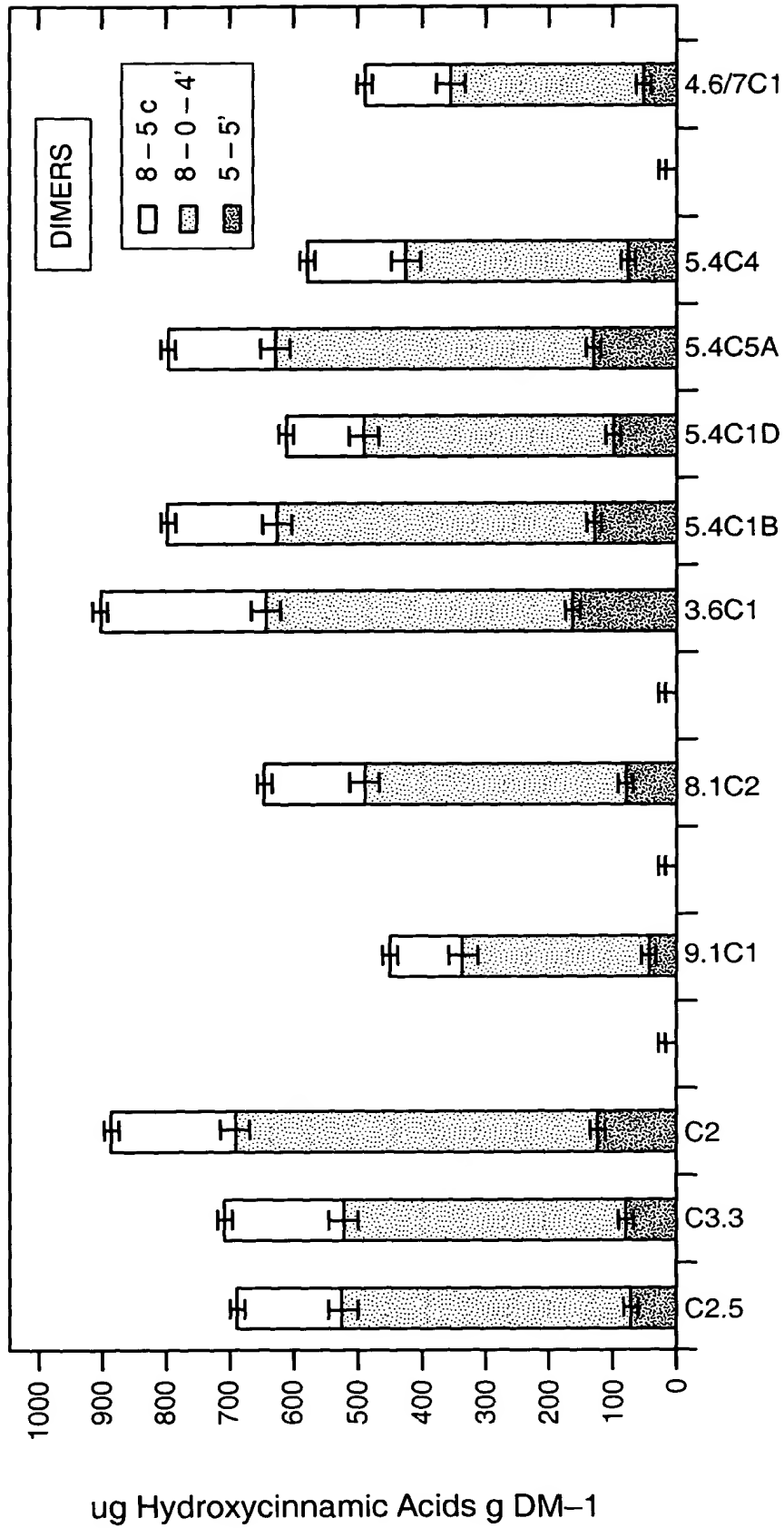


FIG. 22B

26 / 154

In Vitro Dry Matter Digestibility of Leaf Tissue of Mature *Festuca arundinacea* Plants Expressing FAE Under an Actin Promoter

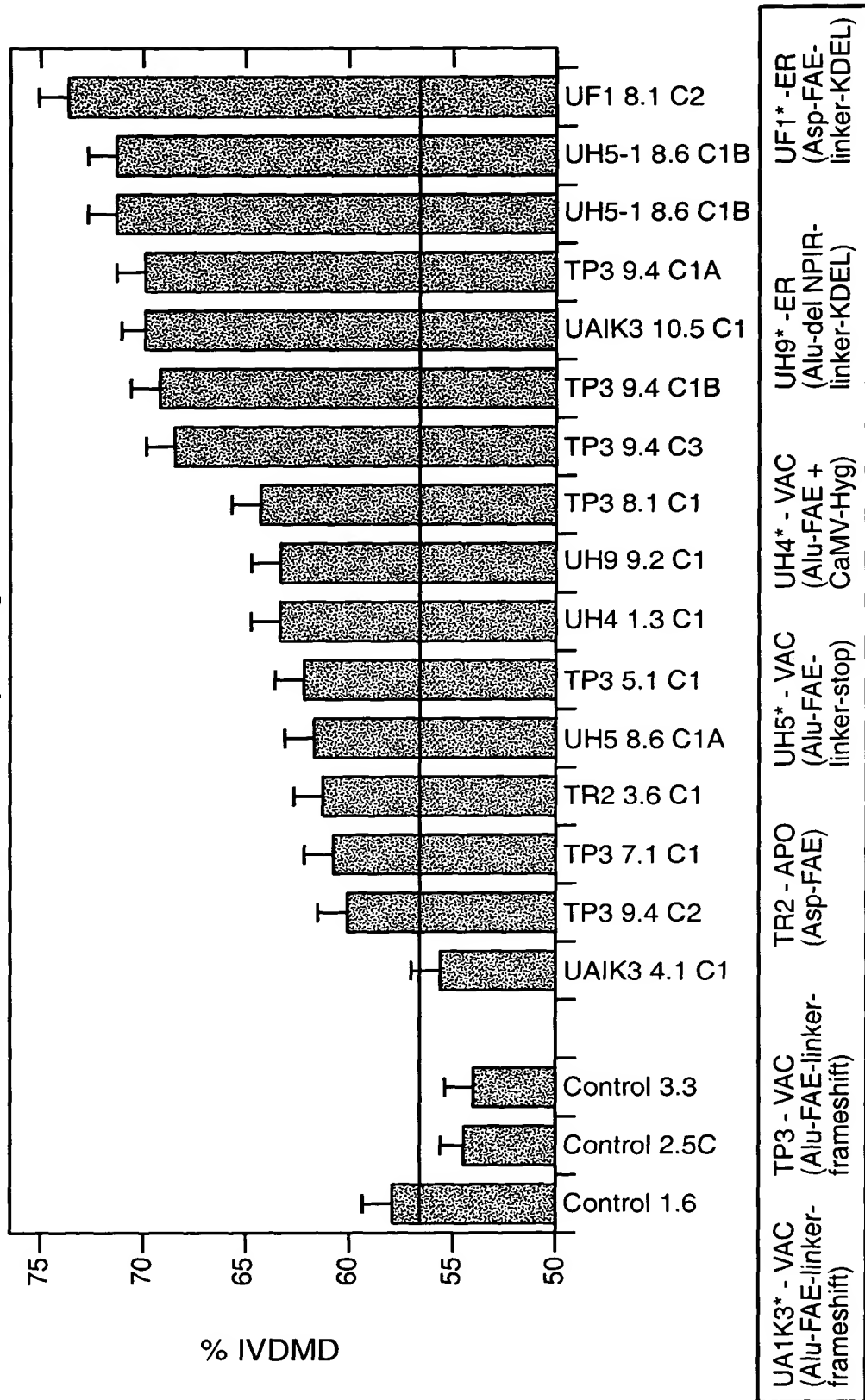


FIG. 23

27 / 154

In Vitro Dry Matter Digestibility of Leaf Tissue of Mature  
*Lolium multiflorum* Plants Expressing FAE Under an Actin Promoter

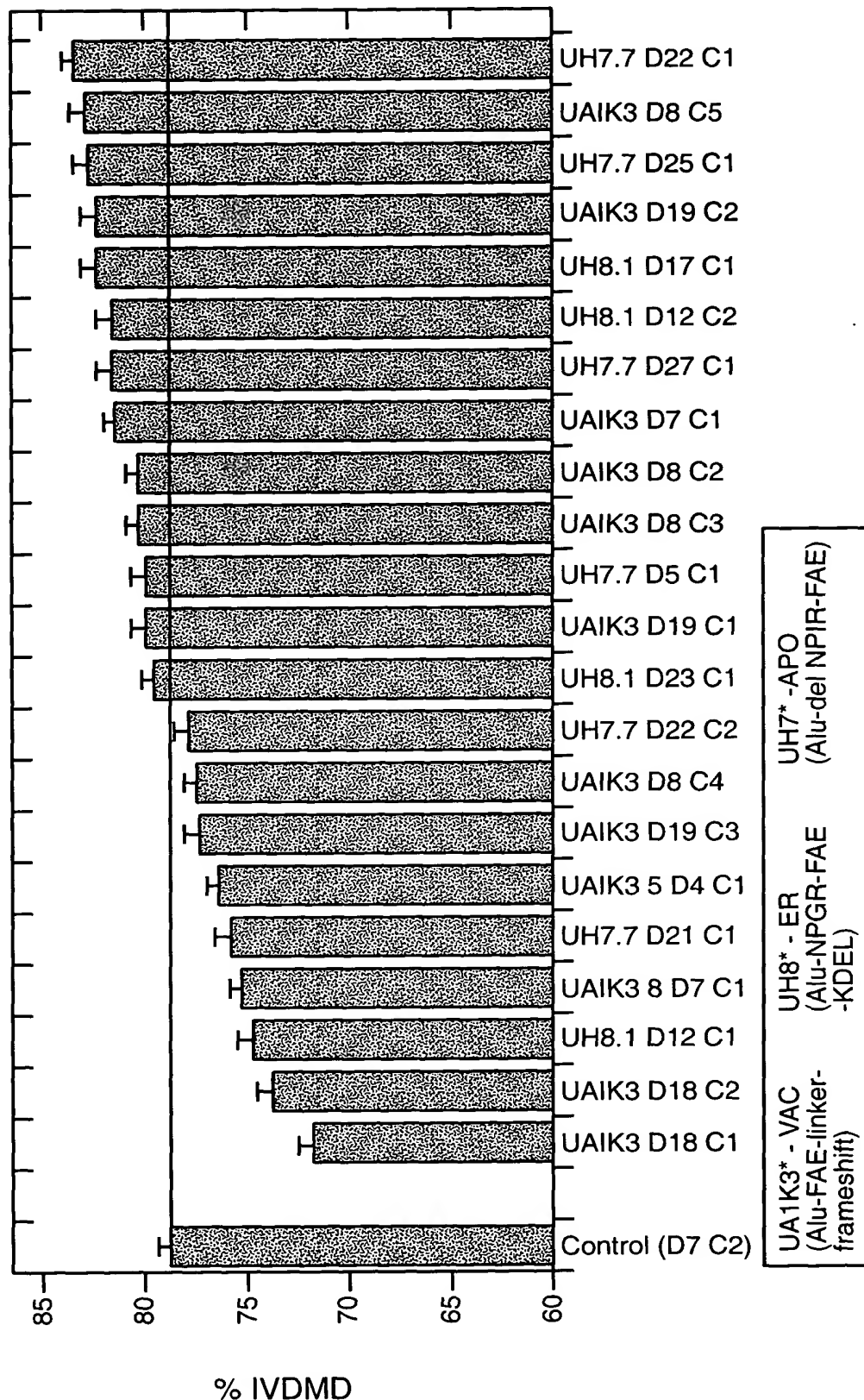


FIG. 24

28 / 154

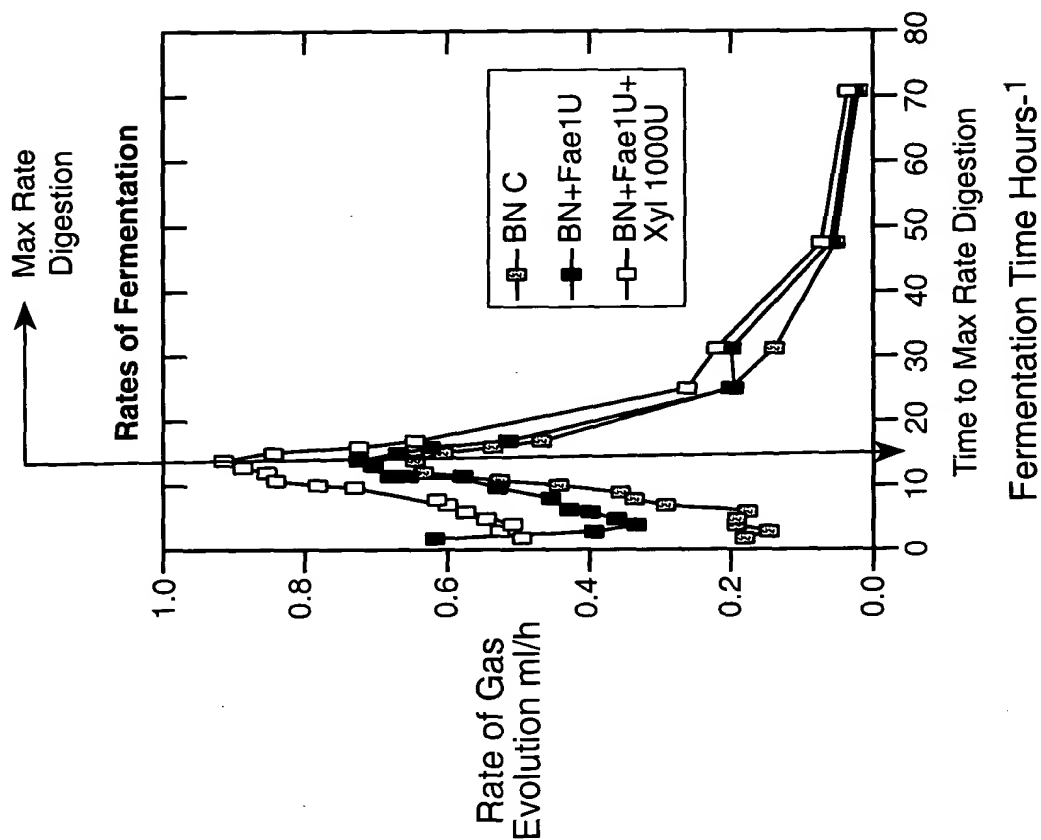


FIG..25B

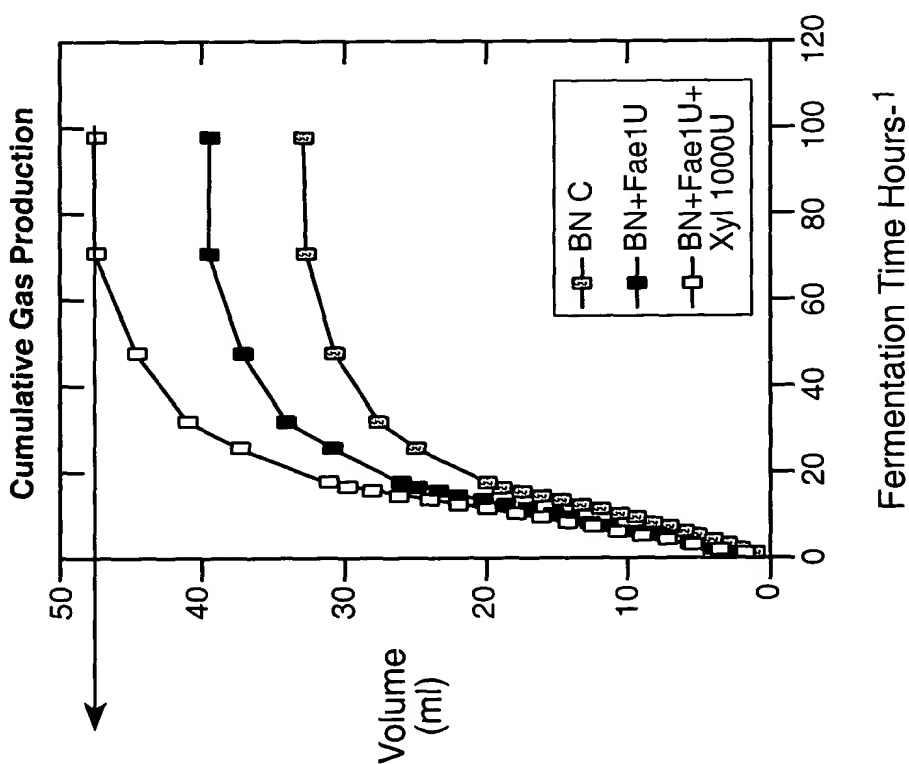
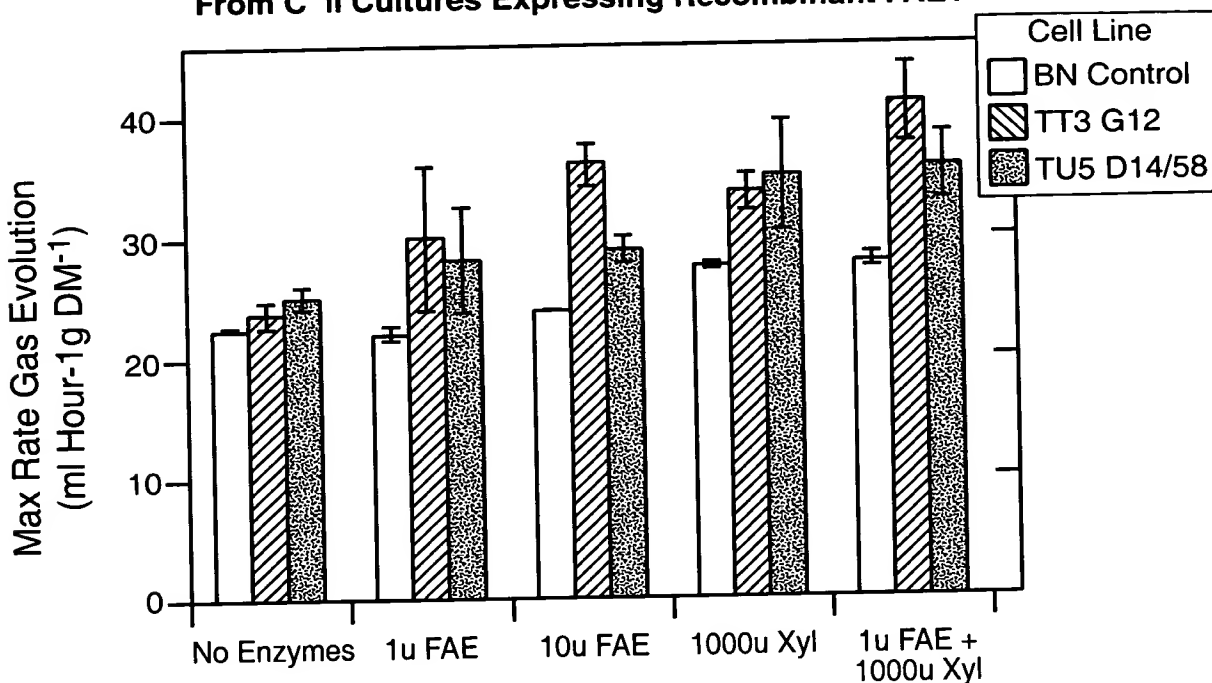


FIG..25A



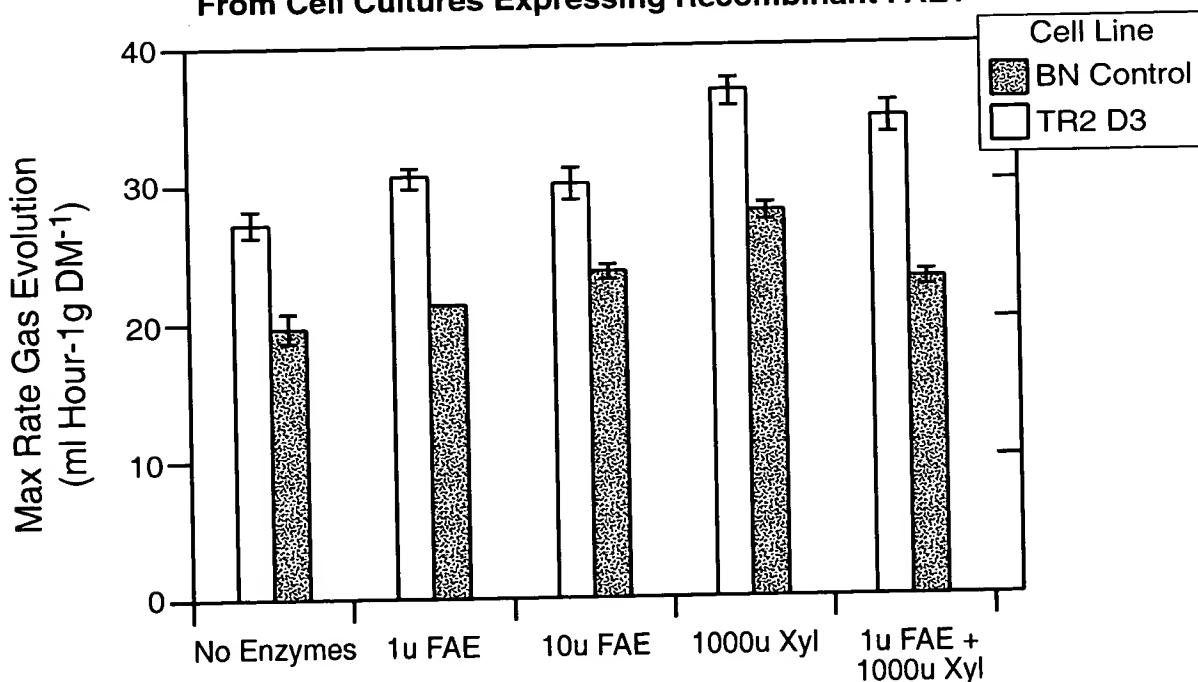
29 / 154

**In-vitro Fermentation of *Festuca arundinacea* C II Walls  
 From C II Cultures Expressing Recombinant FAE1**



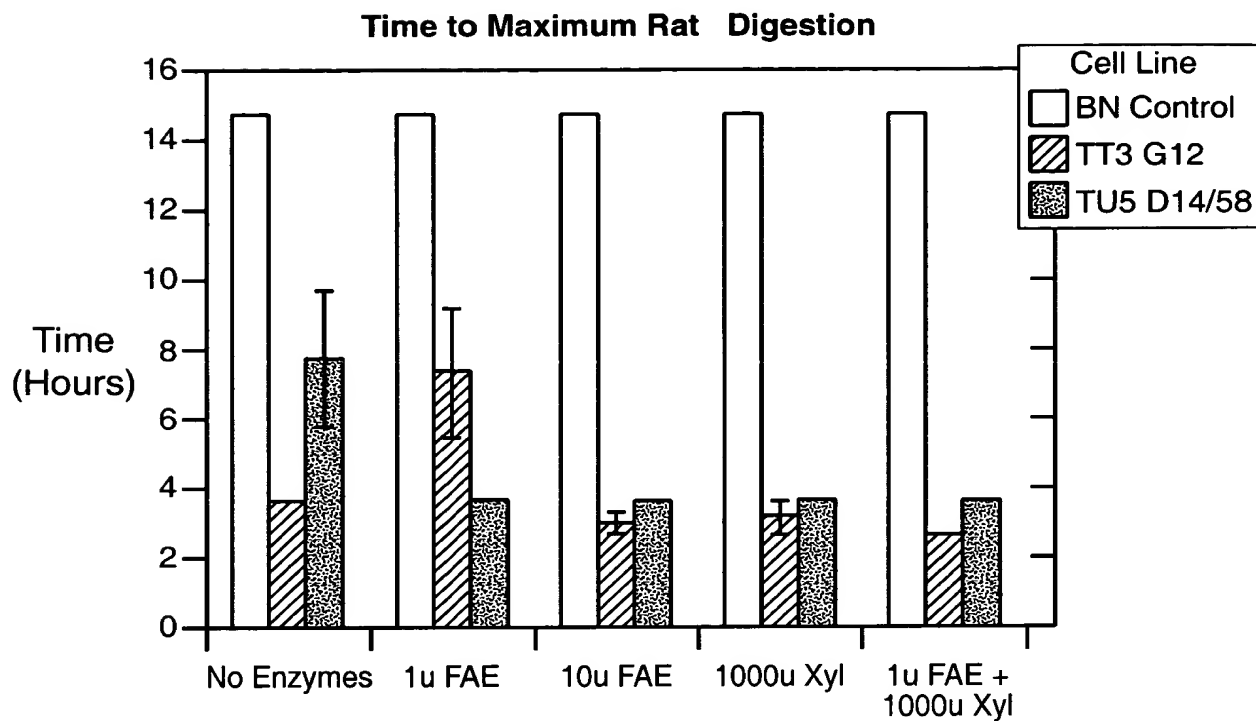
**FIG.\_26A** Maximum Rate of Digestion

**In-vitro Fermentation of *Festuca arundinacea* Cell Walls  
 From Cell Cultures Expressing Recombinant FAE1**

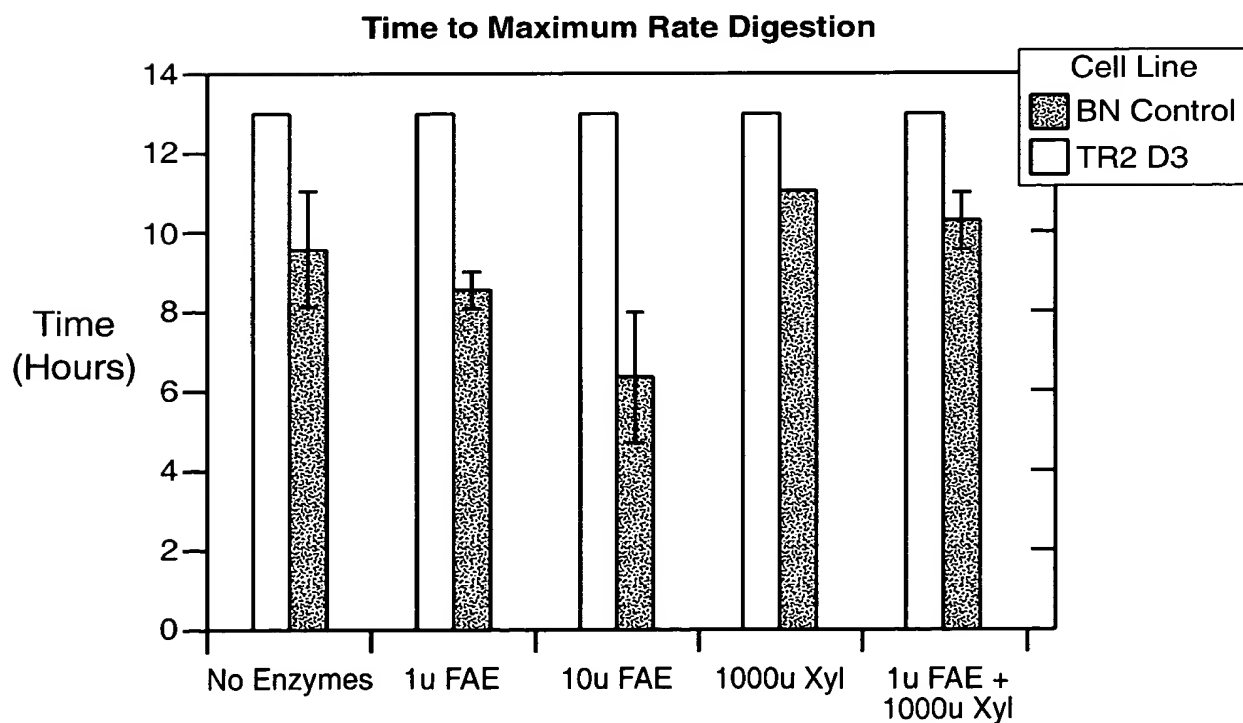


**FIG.\_26B** Maximum Rate of Digestion

30 / 154

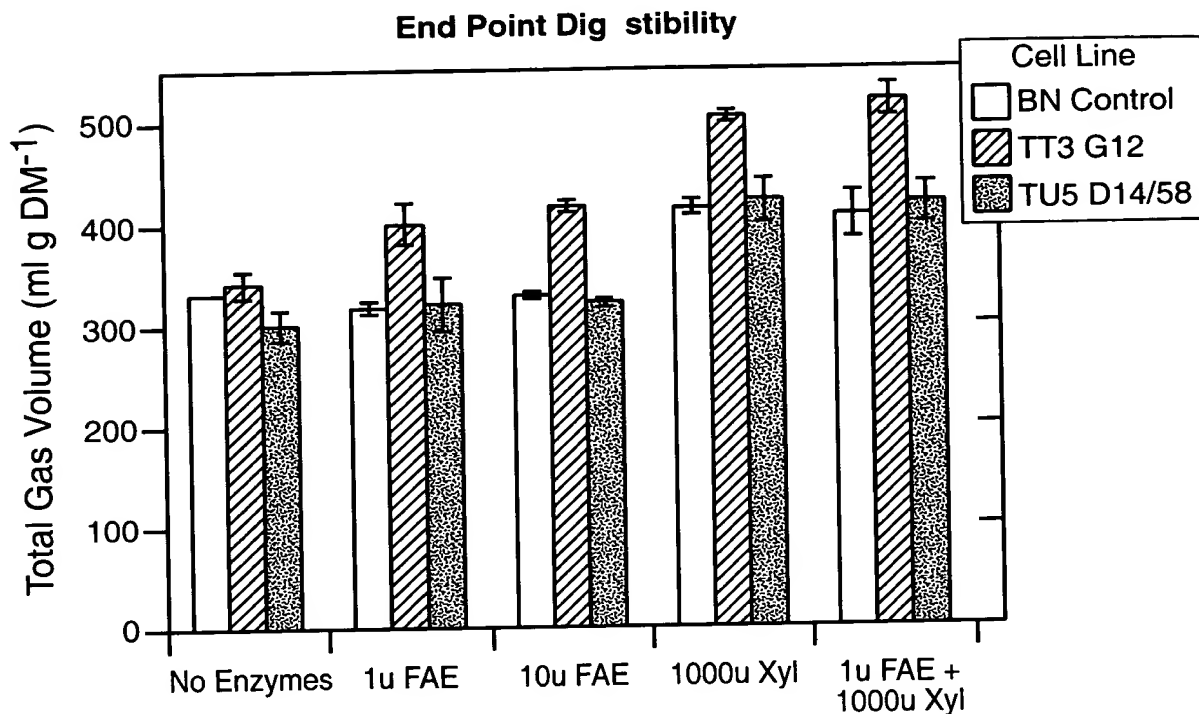


**FIG.\_27A**

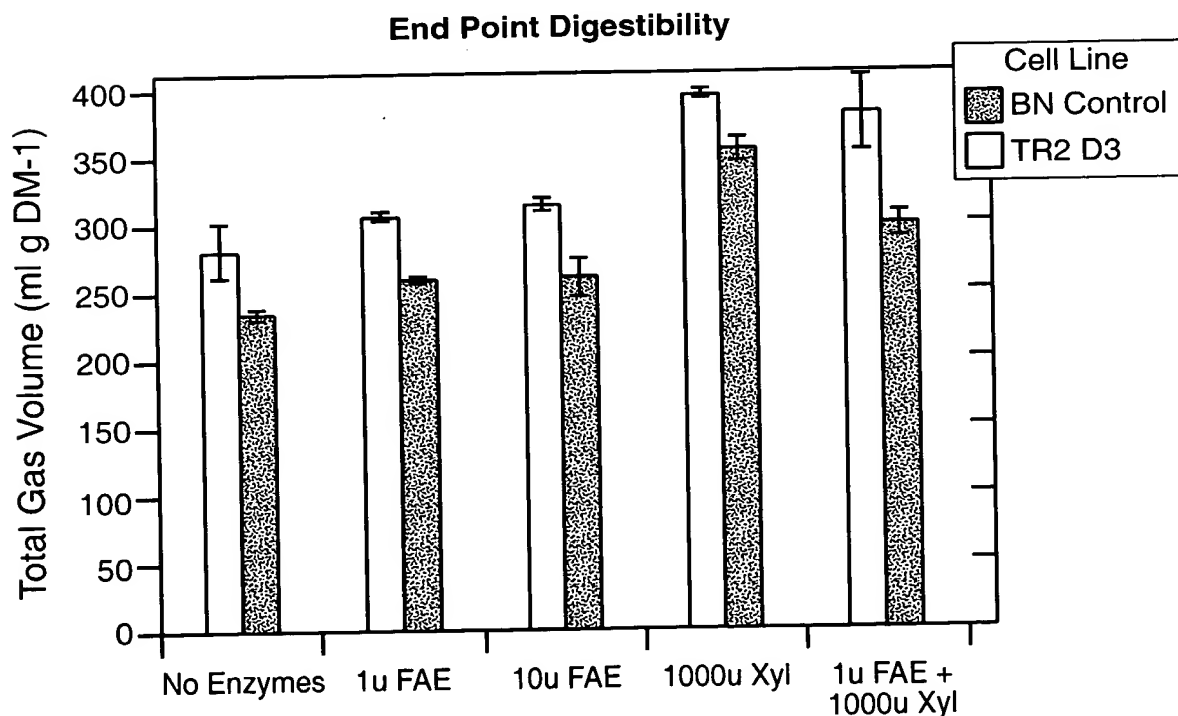


**FIG.\_27B**

31 / 154



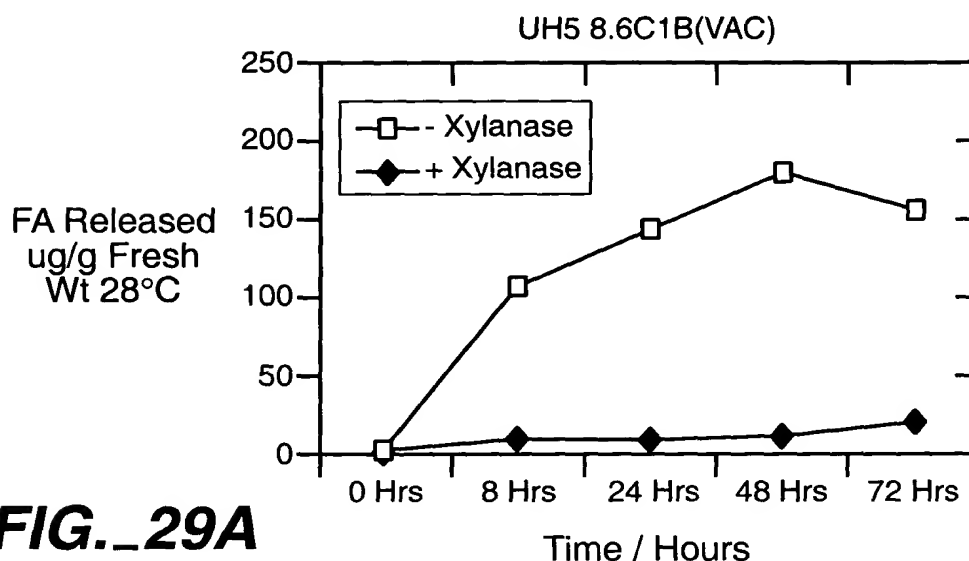
**FIG.\_28A**



**FIG.\_28B**

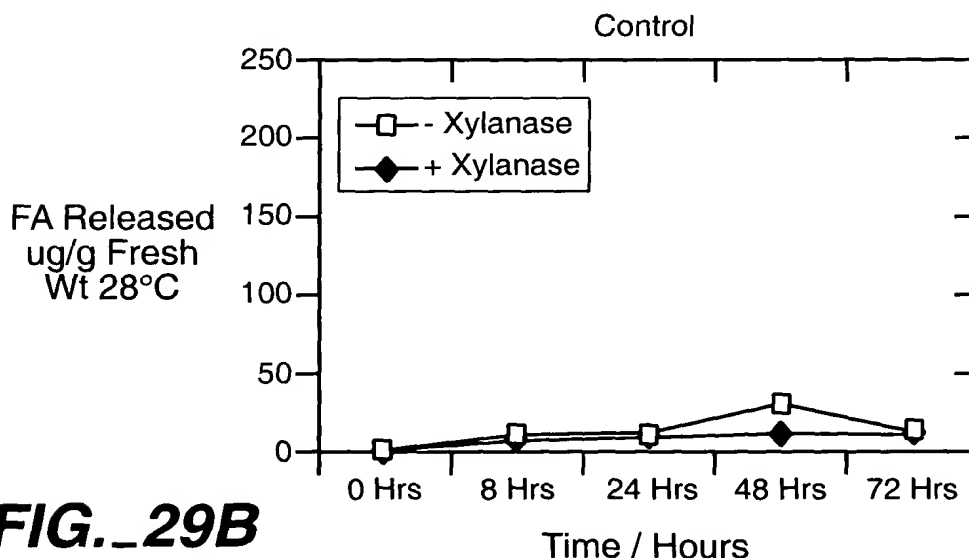
32 / 154

**Kinetics of FAE Activity by Ferulic Acid Release  
 from Cell Wall under Self Digestion in *Festuca arundinacea*  
 and Stimulation by Xylanase**



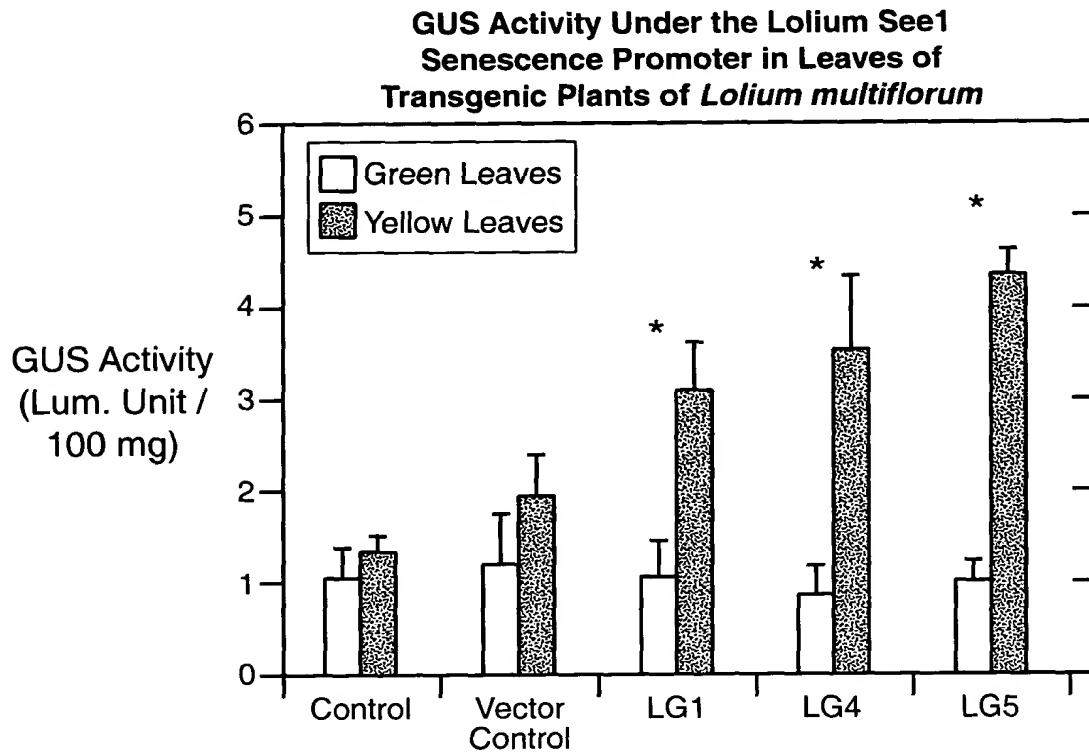
**FIG.\_29A**

**Kinetics of FAE Activity by Ferulic Acid Release  
 from Cell Wall under Self Digestion in *Festuca arundinacea*  
 and Stimulation by Xylanase.**



**FIG.\_29B**

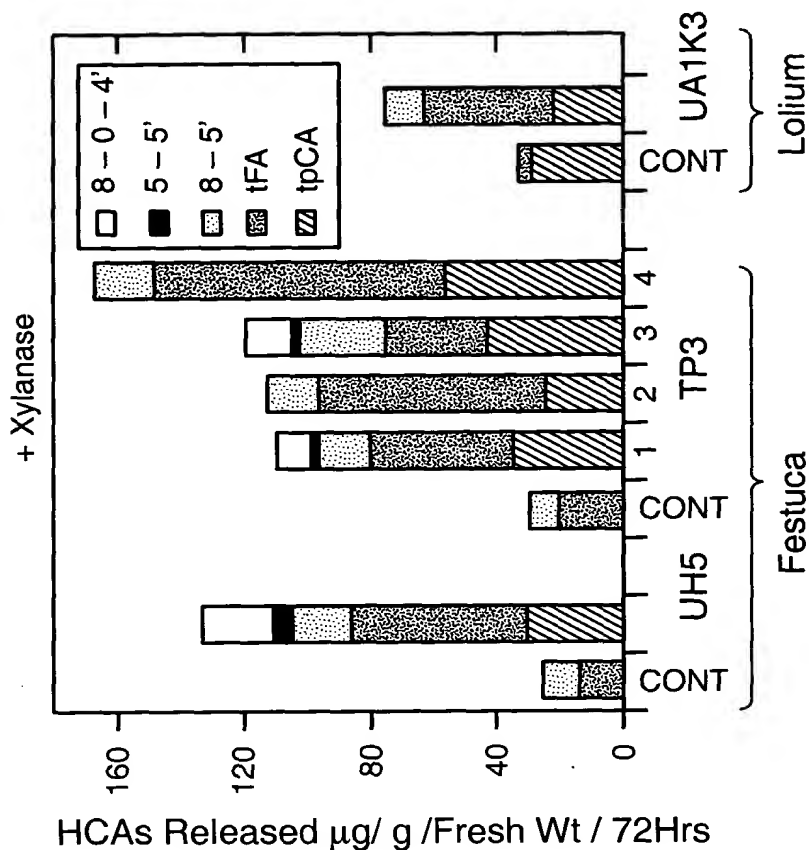
33 / 154



**FIG.\_30**

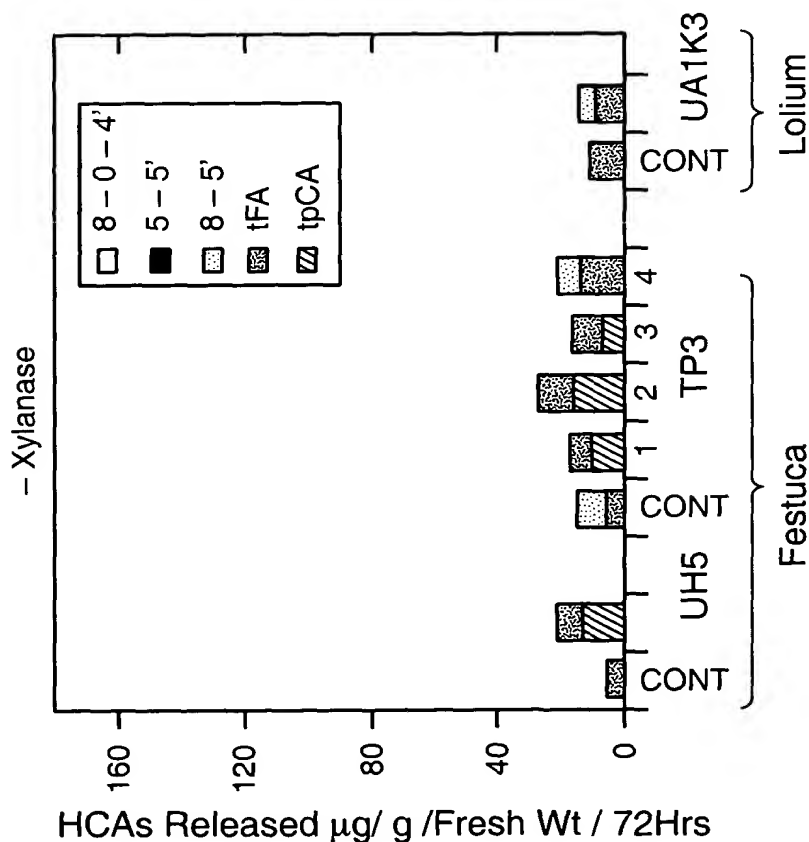
34 / 154

**Release of Monomeric and Dimeric HCAs  
on Self Digestion of Leaves of Vacuolar  
Targeted FAE Expressing Plants**



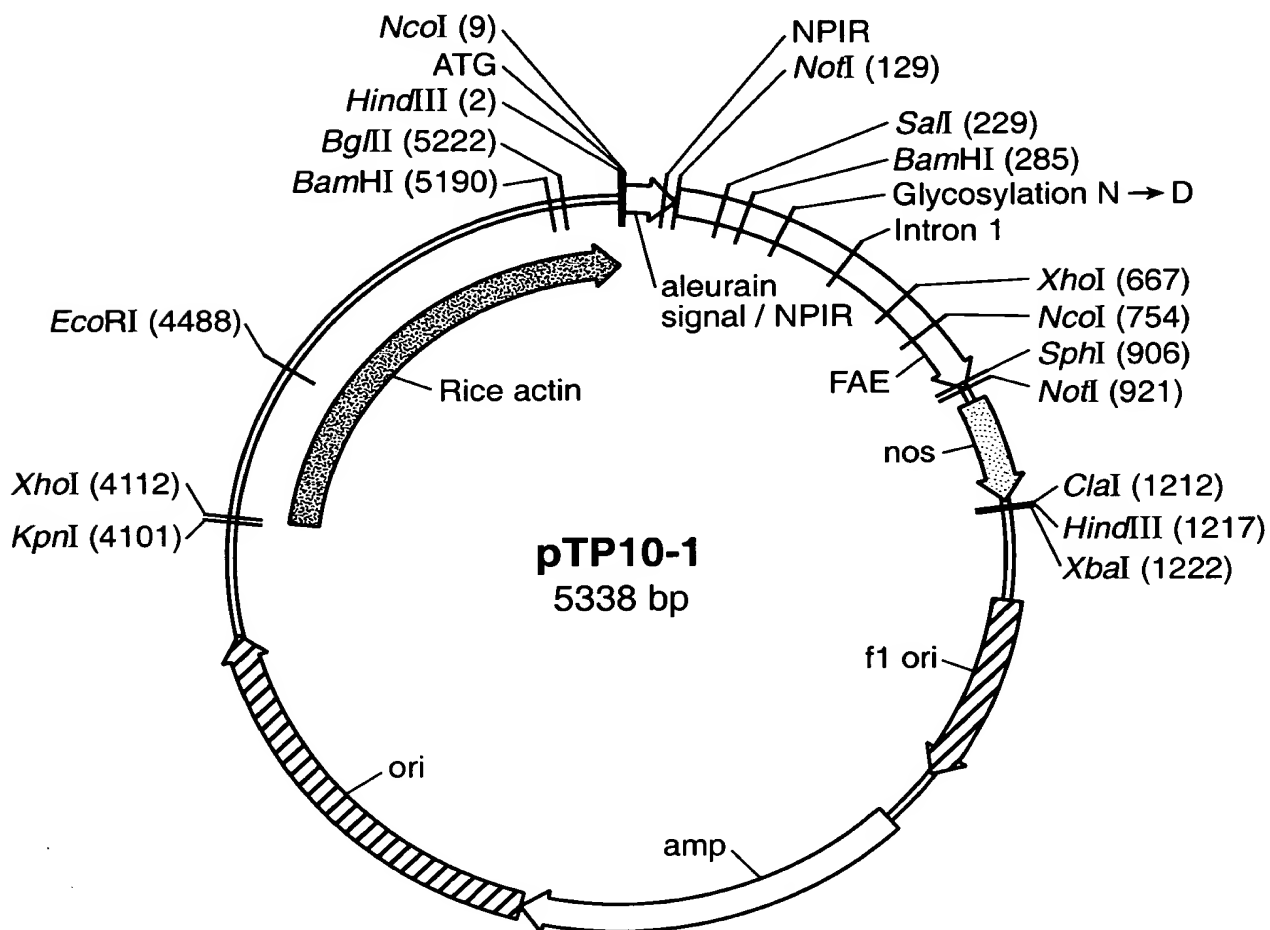
**FIG.\_31B**

**Release of Monomeric and Dimeric HCAs  
on Self Digestion of Leaves of Vacuolar  
Targeted FAE Expressing Plants**



**FIG.\_31A**

35 / 154



**FIG.\_32A**

36 / 154

```

 NcoI
                                ~~~~~
HindIII
~~~~~
 M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACAGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GCCACAGGCC GCCGTCGCCG

 NPIR
                                ~~~~~
                                NotI
                                ~~~~~
 . A S S S F A D S N P I R P V T D R A A A S T .
71 TCGCCTCCTC CTCCTCCTTC GCCACTCCA ACCGATCCG GCCGTCACC GACCGCGCGG CCGCCTCCAC
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC

 Sali
                                ~~~~~
      D   L   C   N   I   P   S   T   I   I   K   G   E   K   I   Y   N   S   Q   T   D   I   N   G
211 GACCTGTGCA ACATCCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAACT GACATTAACG

                                BamHI
                                ~~~~~
 . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTCCGT GGCACCTGGTA GTGATACGAA

 Glycosylation
                                ~~~~~
      .   L   Q   L   D   T   D   Y   T   L   T   P   F   D   T   L   P   Q   C   N   G   C   E   V
351 TCTACAACTC GATACTGACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
      H   G   G   Y   Y   I   G   W   V   S   V   Q   D   Q   V   E   S   L   V   K   Q   Q   V   S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA
      .   Q   Y   P   D   Y   A   L   T   V   T   G   H   X   L   G   A   S   L   A   A   L   T   A   .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
      .   A   Q   L   S   A   T   Y   D   N   I   R   L   Y   T   F   G   E   P   R   S   G   N   Q
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
```

FIG.\_32B



37 / 154

```

XhoI
~~~~~
631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T
 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGGTCA

 NcoI
      ~~~~~
      . H A N D G I P N L P P V E Q G Y A H G G V E Y .
701  CTCATGCCAA CGACGGCATC CCAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA
      . W S V D P Y S A Q N T F V C T G D E V Q C C E
771  CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACCTGGG ATGAAAGTGA GTGCTGTGAG

      SphI
      ~~~~~
841 A Q G G Q G V N N A H T T Y F G M T S G A C T W
 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT

 NotI
      ~~~~~
      . P V A A A E T T E G *
911  GGCCGGTCGC GGCCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAAA CATTTGGCAA
981  TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGCGA TGATTATCAT ATAAATTTCTG TTGAATTACG
1051 TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC
1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAAATATAG CGCGCAAACT AGGATAAATT ATCGCGCGCG

      HindIII
      ~~~~~
 ClaI XbaI
      ~~~~~
1191 GTGTCATCTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTCCG CCTATAGTGA
1261 GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTTACAACGT CGTGACTGGG AAAACCCCTGG CGTTACCCAA
1331 CTTAATCGCC TTGCAGCACA TCCCCCTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC
1401 CTTCCCAACA GTTGGCAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GGCGCATTA GCGCGCGCGG
```

FIG.\_32C

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCCTT CGCTTCTCTC  
 1541 CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCCCTCAAG CTCTAAATCG GGGCTCCCTT TAGGGTTCC  
 1611 GATTAGTGC TTTACGGCAC CTCGACCCCA AAAAATTGA TTAGGGTGAT GGTTCACGTA GTGGGCCATC  
 1681 GCCCTGATAG ACGGTTTTC GCCCTTTGAC TATCTCGGTC TATCTTTTG AATTATAAGG GATTTGCGG ATTTCCGCTT  
 1751 ACTGGAACAA CACTCAACCC CACTCAACCC TATCTCGGTC TATCTTTTG AATTATAAGG GATTTGCGG ATTTCCGCTT  
 1821 ATTTGGTTAA AATGAGCTG AATGAGCTG AATTAACAA AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 1891 TTAGGTGGCA CTTTTCGGG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 1961 TGATCCGCT CATGAGACAA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2031 TCAACATTTT CCGTTCGCTT TATTTCCCTT TATTTCCCTT TATTTCCCTT TATTTCCCTT TATTTCCCTT TATTTCCCTT  
 2101 ACGCTGGTGA AAGTAAAGA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2171 ACAGCGGTAA GATCCTTGAG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2241 GCTATGTGGC GCGGTATTAT AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2311 CAGATGACT TGGTTGAGTA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2381 TATGCAGTGC TGCCATAACC AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2451 GAAGGAGCTA ACCGCTTTT AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2521 CTGAATGAAG CCATACCAA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2591 AACTATTAACT TGGCGAACTA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2661 AGTTGCAGGA CCACTTCTGC AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2731 GAGCGTGGGT CTCGCGGTAT AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2801 ACACGACGGG GAGTCAGGCA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2871 TAAGCATGGG TAACGTGTCAG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 2941 TTTAAAAGGA TCTAGGTGAA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3011 TCCACTGAGC GTCAGACCCC AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3081 CTGCTGCTTG CAACAAAAG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3151 CTTTTCCTGA AGGTAACCTG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3221 TAGGCCACCA CTTCAAGAAC AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3291 TGCTGCCAGT GGCATAAAGT AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3361 CGGTCGGGCT GAACGGGGG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3431 ACCTACAGCG TGAGCTATGA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3501 CGGCAGGGTC GGAACAGGAG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3571 GTCGGGTTTC GCCACCTCTG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3641 AAAACGCCAG CAACGCGGCC AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3711 TGCCTTATCC CCTGATTCTG AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3781 CGAACGACCG AGCGCAGCGA AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT  
 3851 CCGCGCGTTG GCCGATTCTAT AATGAGCTG AATGAGCTG AATTAACGC GAATTTTAA TTTGTTTAA CATTTACAAT

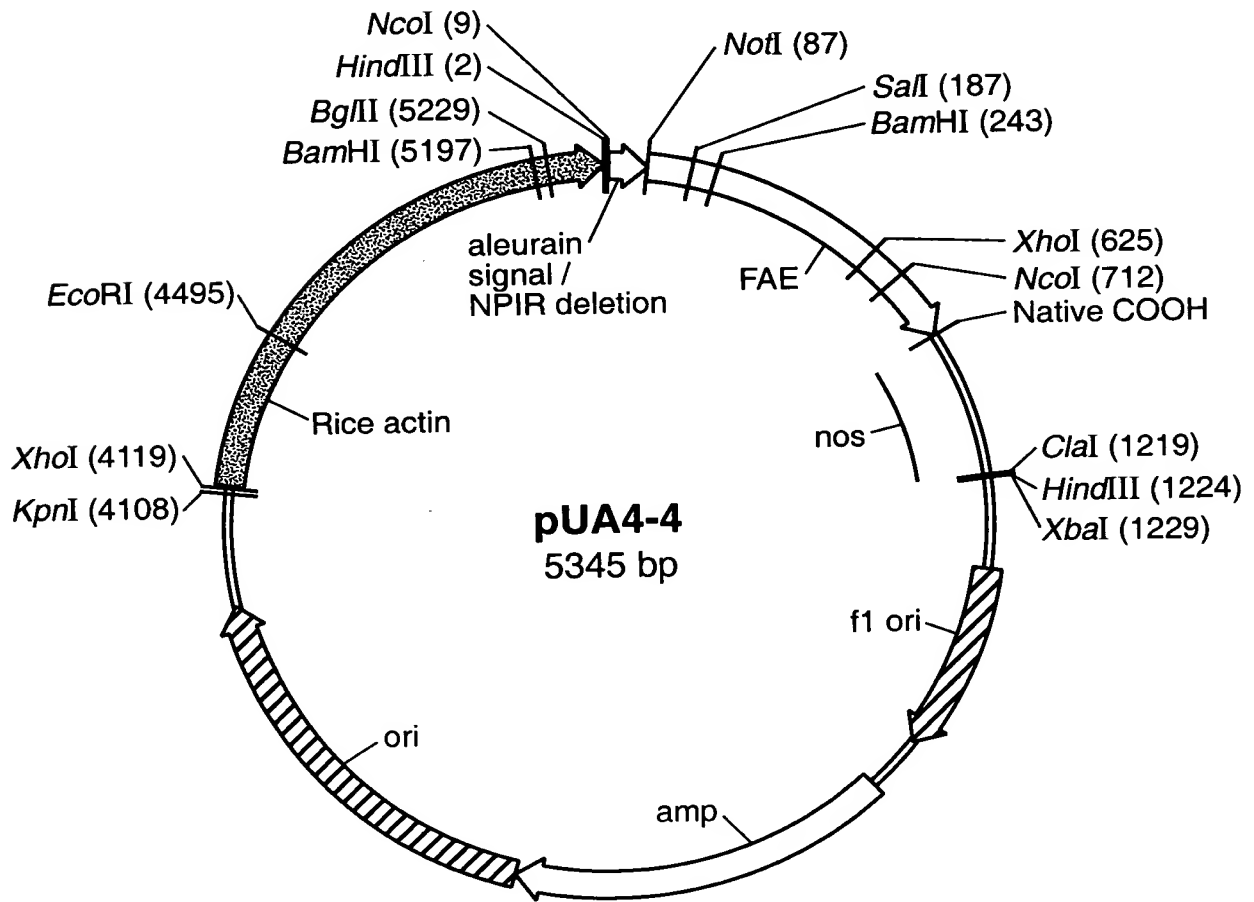
FIG.\_32D

39 / 154

```
3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC
                                     KpnI
4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTCGAGGTCA TTCATATGCT
4131 TGAGAAGAGA GTCGGGATAG TCCAAAATAA AACAAAGGTA AGATTACCTG GTCAAAAGTG AAAACATCAG
4201 TTAAAAGGTG GTATAAGTAA AATATCGGTA ATAAAGGTG GCCCAAAGTG AAATTACTC TTTTCTACTA
4271 TTATAAAAAA TGAGGATGTT TTGTCGGTAC TTTGATACGT CATTTTGTG TGAATTGGTT TTTAAGTTTA
4341 TTCGCGATTT GGAATGCAAT ATCTGTATTT GAGTCGGTTT TTAAGTTCTG TGCTTTTGT AATACAGAGG
4411 GATTGTATA AGAATATCT TTAATAAAC CATATGCTAA TTTGACATAA TTTTGTGAGAA AAATATATAT
                                     EcorI
4481 TCAGGCGAAT TCCACAATGA ACAATAATA GATTAAATA GCTTGCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAAGATA AACTTAGACT CAAAACATTT ACATAAACAA CCCCATAAGT CCTAAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCCC AACCCACCCC AGTGCAGCCA
4691 ACTGGCAAAAT AGTCTCCACC CCGGCACCTA TCACCGTGAG TTGTCCGCAC CACCGCACGT CTCGCAGCCA
4761 AAAAAAATAA AAGAAAGAAA AAAAAGAAA AGAAAACAG CAGGTGGGTC CGGGTCGTGG GGGCCGGA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGG CCGGCCCTCC CTCCGCTTCC AAAGAAACGC CCCCCATCGC
4901 CACTATATAC ATACCCCCC CTCTCTCTCC ATCCCCCCTA CCTACACCA ACCACCACTC
4971 CCCCCCTCGT GCCGGACGAC GAGCTCCTCC CCCCCTCCCC TCCGCCGCCG CCGGTAACCA CCCCCCCC
5041 CTCCTCTTTC TTTCTCCGTT TTTTCTTCG TCTCGGTC TCCTCGGTC GATCTTTGGC CTGGTAGTT TGGGTGGCG
5111 AGAGCGGCTT CGTCGCCCG CAGATCGGCG GAGAGGGCG BglII
                                     BglII
5181 GTCGGCCCCG ATCCTCGCGG GGAATGGGG TCCTCGGATGT AGATCTTCTT TCCTTCTTCT TTTTGTGGTA
5251 GAATTTGAAT CCTCAGCAT TGTTCATCGG TAGTTTTTCT TTTTCATGAT TTGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTTGTAGC
```

FIG.\_32E

40 / 154



**FIG.\_33A**

41 / 154

```

                                NcoI
                                ~~~~~
HindIII
~~~~~
      M   A   H   A   R   V   L   L   L   A   L   A   V   L   A   T   A   A   V   A   V
1  AAGCTTACCA TGGCCACAGC CCGCGTCTCTC CTCTTGCGGC TCGCCGTGCT GCCACAGGCC GCCGTGCGCG
                                NotI
                                ~~~~~
 . A S S R A A A S T Q G I S E D L Y S R L V E M .
71 TCGCCTCCTC CCGCGCGGCC GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT
 Sali
                                ~~~~~
      .   A   T   I   S   Q   A   A   Y   A   D   L   C   N   I   P   S   T   I   I   K   G   E   K
141  GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTGCA CTATTATCAA GGGAGAGAAA
                                BamHI
                                ~~~~~
 I Y N S Q T D I N G W I L R D D S S K E I I T V
211 ATTTACAATT CTCAAACTGA CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAATCACCG
 . F R G T G S D T N L Q L D T N Y T L T P F D T .
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACTCGA TACTAACTAC ACCCTCACGC CTTTCGACAC
 . L P Q C N G C E V H G G Y Y I G W V S V Q D Q
351 CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA
 V E S L V K Q Q V S Q Y P D Y A L T V T G H X L
421 GTCGAGTCGC TTGTCAAACA GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCCACKCCC
 . G A S L A A L T A A Q L S A T Y D N I R L Y T .
491 TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TCGGACATAC GACAACATCC GCCTGTACAC
 XhoI
                                ~~~~~
      .   F   G   E   P   R   S   G   N   Q   A   F   A   S   Y   M   N   D   A   F   Q   A   S   S
561  CTTCCGGCGAA CCGCGCAGCG GCAATCAGGC CTTCGCGTCG TACATGAACG ATGCCCTTCCA AGCCTCGAGC
      P   D   T   T   Q   Y   F   R   V   T   H   A   N   D   G   I   P   N   L   P   P   V   E   Q
631  CCAGATACGA CGCAGTATTT CCGGCTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC CCGGTGGAGC
                                NcoI
                                ~~~~~
 . G Y A H G G V E Y W S V D P Y S A Q N T F V C .

```

FIG.\_33B

```

701 AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACA CATTGTCTCG
 . T G D E V Q C C E A Q G G Q G V N A H T T Y
771 CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGTGTGA ATAATGCGCA CACGACTTAT
 F G M T S G A C T W *
841 TTTGGGATGA CGAGCGGAGC CTGTACATGG TGATCAGTCA TTTCAGCCTC CCCGAGTGTA CCAGGAAAGA
911 TGGATGTCCT GGAGAGGGGG CCGCGTAACC ACTGAAGGAT GAGCTGTAAA GAAAGCAGATC GTTCAAAACAT
981 TTGGCAATAA AGTTTCTTAA GATTGAATCC TGTTGCCGGT CTTGCCGATGA TTATCATATA ATTCTGTGTTG
1051 AATTACGTTA AGCATGTAAT AATTAAACATG TAATGCATGA CGTTATTTAT GAGATGGGTT TTTATGATTA
1121 GAGTCCCCGA ATTATACATT TAATACGCGA TAGAAAACAA AATATAGCGC GCAAACTAGG ATAAATTATC

 HindIII
                                ~~~~~
                                ClaI      XbaI
                                ~~~~~
1191 GCGCGCGGTG TCATCTATGT TACTAGATCG ATAAGCTTCT AGAGCGGCGG GTGGAGCTCC AATTCGCCCT
1261 ATAGTGAGTC GTATTACGCG CGCTCACTGG CCGTCGTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT
1331 TACCCAACTT AATCGCCTTG CAGCACATCC CCTTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACC
1401 GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGGACGCGCC CTGTAGCGGC GCATTAAGCG
1471 CGGCGGGTGT GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAAGGCC CTAGCGCCCG CTCCTTTTCG
1541 TTTCTTCCCT TCCTTTCTCG CCACGTTCCG CGGCTTTCCC CGTCAAGCTC TAAATCGGGG GCTCCCTTTA
1611 GGGTTCGGAT TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAAGT
1681 GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCCTTTAATA GTGGACTCTT
1751 GTTCCAAACT GGAACAAAC GAAACCCAT TCAACCCAT CTCGGTCTAT TCTTTTGAT TATAAGGGAT TTTGCCGATT
1821 TCGGCCTATT GGTAAACAAA TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAAACAAA ATATTAACGC
1891 TTACAATTTA GTGGCACTT TTGCGGGAAT TGTCGCGGGA ACCCTATTT GTTTATTTT CTAAATACAT
1961 TCAATATATG ATCCGCTCAT GAGACAATAA CCTGTATAA TGCTTCAATA ATATGAAAA AGGAAGAGTA
2031 TGAGTATTCA ACATTTCCGT GTCGCCCTTA TTCCCTTTT TCGCGCATTT TGCCCTCCCTG TTTTGTGCTCA
2101 CCCAGAAACG CTGGTGAAG TAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTA CATCGAACTG
2171 GATCTCAACA GCGGTAAGAT CTTTGAGAGT TTTTCGCCCG AAGAACGTTT TCCAATGATG AGCACTTTTA
2241 AAGTTCTGCT ATGTGGCGG GTATTATCCC GTATTGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA
2311 CTATTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAGCATC TTACGGGATG CATGACAGTA
2381 AGAGAATTAT GCAGTGCTGC CATAACCATG AGTATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG
2451 GAGGACCGAA GGAGCTAACC GCTTTTTCG ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA
2521 ACCGGAGCTG AATGAAGCCA TACCAAACGA CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAACG
2591 TTGCGCAAAC TATTAACCTG CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG
2661 CGGATAAAGT TGCAGGACCA CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTTATTGCTG ATAAATCTTG

```

FIG.\_33C

```

2731 AGCCGGTGAG CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAGGCCCTC CCGTATCGTA
2801 GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT
2871 CACTGATTAA GCATTGGTAA CTGTCAGACC AAGTTTACTC ATATATACTT TAGATTGATT TAAACATTCA
2941 TTTTAAATTT AAAAGGATCT AGGTGAAGAT CCTTTTGTAT AATCTCATGA CCATAATCCC TTAACGTGAG
3011 TTTTCGTTC ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT TTTTCTCTGC
3081 GCGTAATCTG CTGCTTGCAA ACAAAAAC CACCGTACC AGCGTGGTT TGTTGCCGG ATCAAGAGCT
3151 ACCAACTCTT TTTCCGAAGG TAACTGGCTT GATAGACCGC CAGATACCAA ATACTGTCTT TCTAGTGTAG
3221 CCGTAGTTAG GCCACCACTT TCCAGTGGC GATAAGTCTT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTGTAC
3291 CAGTGGCTGC TGCCAGTGGC TCGGGCTGAA GCTATGAGAA AGCGCCACAG CCCAGCTGG AGCGAACGAG CTACACCGAA
3361 GCGCAGCGG TACAGCGTGA CAGGTCGGA CAGGTCGGA GCGCCACGC GTCCCGAAG TTCCCGAAG GAGAAAGCGG GACAGGTATC
3431 CTGAGATACC TACAGCGTGA CAGGTCGGA CAGGTCGGA GCGCCACGC GTCCCGAAG TTCCCGAAG GAGAAAGCGG GACAGGTATC
3501 CCGTAAGCGG CAGGTCGGA CAGGTCGGA CAGGTCGGA GCGCCACGC GTCCCGAAG TTCCCGAAG GAGAAAGCGG GACAGGTATC
3571 TAGTCCTGTC GGTTCGCGC ACCTCTGACT TGAGCGTCCA TTTTGTGAT TTTTGTGAT GCTCGTCAGG GGGCGGGAGC
3641 CTATGGAAA ACGCCAGCAA GGTATCCCTT GATTCTGTGG ATACCCTTTT TACCGCTTTT CTGGCCTTTT GCTCACATGT
3711 TCTTCTCTGC GTTATCCCTT GATTCTGTGG ATACCCTTTT TACCGCTTTT TACCGCTTTT GAGTGAGCTG ATACCCTTTT
3781 CCGCAGCCGA ACGACCGAGC ACGACCGAGC ACGACCGAGC ACGACCGAGC ACGACCGAGC ACGACCGAGC ACGACCGAGC
3851 CCTCTCCCCG CGCGTTGGCC GATTCTGTAA TGACGCTGGC TGACGCTGGC TGACGCTGGC TGACGCTGGC TGACGCTGGC
3921 GTGAGCGCAA CGCAATTAAT GTGAGTTAGC TCACTCATTA GGCACCCCGG GCTTTACACT TTAATGCTTCC
3991 GGCTCGTATG TTGTGTGGAA TTGTGTGGCG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC

4061 GCCAAGCGCG CAATTAAACC TCACATAAAG GAACAAAAGC TGGGTACCGG GCCCCCCCTC GAGGTCAATC
4131 ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAATC AAAGGTAAAG TTAACCTGGTC AAAAGTGAAA
4201 ACATCAGTTA AAAGGTGGTA TAAAGTAAAT ATCGGTAATC AAAGGTGGCC CAAAGTGAAA TTTACTCTTT
4271 TCTACTATTA TAAAAATTGA GGATGTTTTG TCGGTACTTT GATACGTCTT TTTTGTATGA ATTGGTTTTT
4341 AAGTTTATTC GCGATTGGA AATGCATATC TGTATTTGAG TCGGTTTTTA AGTTCGTTGC TTTTGTAAAT
4411 ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAACCCAT ATGCTAATTT GACATAAATTT TTGAGAAAAA

4481 TATATATTCA GCGGAATTCC ACAATGAACA ATAATAAGAT TAAAATAGCT TGCCCCCGTT GCAGCGATGG
4551 GTATTTTTC TAGTAAATA AAAGATAAAT TTAGACTCAA AACATTTACA AAAACAACCC CTAAAGTCCT
4621 AAAGCCCAA GTGCTATGCA CGATCCATAG CAAGCCAGC CCAACCCAA CCAACCCAA CCAACCCAA
4691 GCAGCCAACT GGCAAAATAG CTCCACCCCC GGCACATATCA CCGTGAGTTG TCCGACCCAC CGCACGTCCTC
4761 GCAGCCCAA AAAAAAAAG AAAAAAAAG AAAAAAAAG AAAAACAGC GTGGTCCCG GTCGTGGGGG
4831 CCGGAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC CGCTTCCAAA GAAACGCCCC

```

FIG..33D

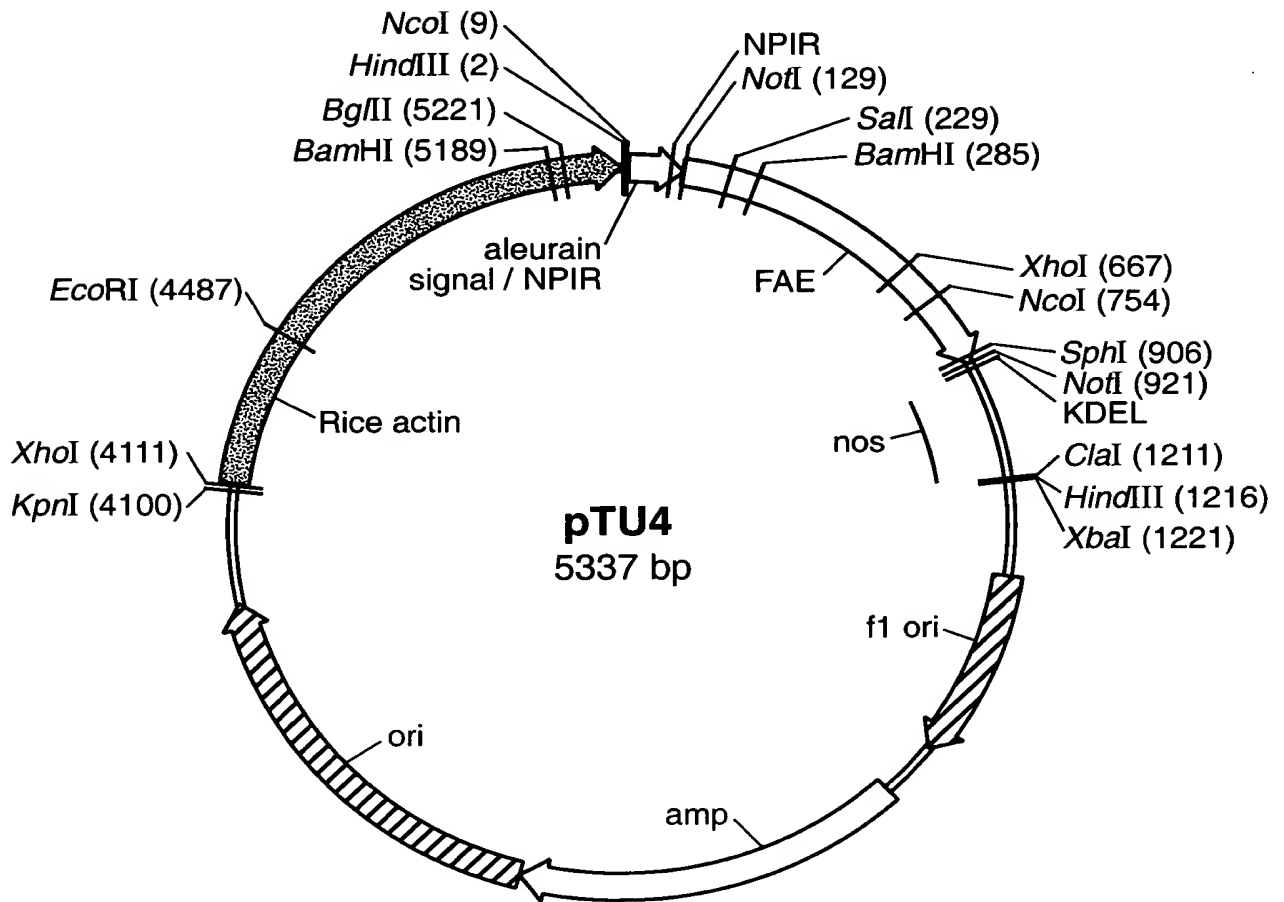
44 / 154

```
4901 CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCACACCC TACCACCACC ACCACCACCA
4971 CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCTCTCCCC CTCCCCCTCC GCCGCCGCCG GTAACCAACCC
5041 CGCCCTCTC CTCTTTCTTT CTCCGTTTTT TTTTTCGTCT CCGTCTCGAT CTTTGGCCCTT GGTAGTTTGG
5111 GTGGGCGAGA GCGGCTTCGT CGCCAGATC GGTGCGCGG GGGGCGGGA TCTCGCGGCT GCGTCTCTCCG
 BamHI
                                     ~~~~~
5181 GCGGTGAGTC GGCCCGGATC CTCGCGGGGA ATGGGCTCT CGGATGTAGA TCTTCTTTCT TTCTTCTTTT
5251 TGTGGTAGAA TTTGAATCCC TCAGCATTGT TCATCGGTAG TTTTCTTTT CATGATTGT GACAAATGCA
5321 GCCTCGTGG GAGCTTTT GTAGC
```

FIG.\_33E



45 / 154



**FIG. 34A**

```

      NcoI
      ~~~~~
HindIII
~~~~~
1  AAGCTTACCA TGGCCCACGC CCGCGTCCCTC CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCCGTCGCCG
      NotI
      ~~~~~
 . A S S S S F A D S N P I R P V T D R A A A S T .
71 TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCGATCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTATAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
 Sali
      ~~~~~
      D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAAAC T GACATTAAACG
      BamHI
      ~~~~~
 . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACCTGGTA GTGATACGAA
 . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAATC GATACTAACT ACACCTCACC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
 . Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GGAATACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTCGGCG AACCGCGCAG CGGCAATCAG
 XhoI
      ~~~~~
      A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGGTCA
      NcoI
      ~~~~~
 . H A N D G I P N L P P V E Q G Y A H G G V E Y .

```

FIG.\_34B

```

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCAGGTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA
 . W S V D P Y S A Q N T F V C T G D E V Q C C E
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACCTGGG ATGAAGTGCA GTGCTGTGAG
 SphI
    ~~~~
    A Q G G Q G V N N A H T T Y F G M T S G A C T W
841 GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGG GCATGCACCT

    NotI
    ~~~~~~
 . P V A A A E P L K D E L *
911 GGCCGGTCGC GGCCGCGGAA CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT
981 AAAGTTTCTT AAGATTGAAT CCTGTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT
1051 TAAGCATGTA ATAATTACAA TGTAAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG
1121 CAATTATACA TTTAATACGC GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG

 HindIII
    ~~~~~~
    ClaI          XbaI
    ~~~~~~
1191 TGTCACTCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC CTATAGTGAG
1261 TCGTATTACG CGGCTCACT GGCCGTCGTT TTACAAACGTC GTGACTGGGA AAACCTTGGC GTTACCCAAAC
1331 TTAATCGCCT TGCAGCACAT CCCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCGCA CCGATCGCCC
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGACGCG CCCTGTAGCG GCGCATTAAG CCGCGCGGGT
1471 GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTCTTCC
1541 CTTCCTTTCT CGCCACGTTT TACGGGCACC TCGACCCCAA AAACTTGAT TAGGGTCCCTT TAGGGTTCCG
1611 ATTTAGTGCT TTACGGCACC CGGTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCCTTAA TAGTGGACTC TTGTTCCAA
1681 CCCTGATAGA CGGTTTTCG ACTCAACCTT ATCTCGGTCT ATCTTTTGA TTTATAAGGG AATTTTAAAC GCTTACAAT
1751 CTGGAACAAC ACTCAACCTT ATCTCGGTCT ATCTTTTGA TTTATAAGGG AATTTTAAAC GCTTACAAT
1821 TTGGTTAAA AATGAGCTGA TTAAACAAA ATTTAACGCG AATTTTAAAC AATATATTAAC GCTTACAAT
1891 TAGGTGGCAC TTTTCGGGA AATGTGCGG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT
1961 GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
2031 CAACATTTC GTGTCGCCCT TATTCCTTTT TTTGCGGCAT TTTGCTTCC ACAGTGGGT TACATCGAAC TGGATCTCAA
2101 CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TTTCCCAATGA TGAGCACTTT TAAAGTTCTG
2171 CAGCGGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT TTTCCCAATGA TGAGCACTTT TAAAGTTCTG
2241 CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC
2311 AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAAAT
2381 ATGCAGTGCT GCCATAACCA TGAGTGATTA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG
```

FIG.\_34C

48 / 154

|      |             |             |             |             |             |             |             |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 2451 | AAGGAGCTAA  | CCGCTTTTTT  | GCACAACATG  | GGGATCATG   | TAACTCGCCT  | TGATCGTTGG  | GAACCGGAGC  |
| 2521 | TGAATGAAGC  | CATACCAAAC  | GACGAGCGTG  | ACACCACGAT  | GCCTGTAGCA  | ATGGCAACAA  | CGTTGCGCAA  |
| 2591 | ACTATTAACT  | GGCGAACTAC  | TTACTCTAGC  | TTCCCGGCAA  | CAATTAATAG  | ACTGGATGGA  | GGCGGATAAA  |
| 2661 | GTTGCAGGAC  | CACCTCTGCG  | CTCGGCCCTT  | CGGCTGGCT   | GGTTTATTGC  | TGATAAATCT  | GGAGCCGGTG  |
| 2731 | AGCGTGGGTC  | TCGGGTATC   | ATTGCAGCAC  | TGGGGCCAGA  | TGGTAAGCC   | TCCCGTATCG  | TAGTTATCTA  |
| 2801 | CACGACGGGG  | AGTCAGGCAA  | CTATGGATGA  | ACGAATAGA   | CAGATCGCTG  | AGATAAGTGC  | CTCACTGATT  |
| 2871 | AAGCATTTGT  | AACGTGTCAGA | CCAAAGTTTAC | TCATATATAC  | TTTAGATTGA  | TTTAAAACTT  | CATTTTTAAT  |
| 2941 | TTAAAAGGAT  | CTAGGTGAAG  | ATCCTTTTGT  | ATAATCTCAT  | GACCAAAATC  | CCTTAAACGTG | AGTTTTCGTT  |
| 3011 | CCACTGAGCG  | TCAGACCCCG  | TAGAAAAGAT  | CAAAAGGATCT | TCTTGAGATC  | CTTTTTCCT   | GCGCGTAAATC |
| 3081 | TGCTGCTTGC  | AAACAAAAAA  | ACCACCGCTA  | CCAGCGGTGG  | TTTGTTTTGC  | GGATCAAGAG  | CTACCAACTC  |
| 3151 | TTTTTCCGAA  | GGTAACCTGGC | TTCAGCAGAG  | CGCAGATACC  | AAATACTGTC  | CTTCTAGTGT  | AGCCGTAGTT  |
| 3221 | AGGCCACCAC  | TTCAAGAACT  | CTGTAGCAC   | GCCTACATAC  | CTCGCTCTGC  | TAATCCCTGTT | ACCAGTGGCT  |
| 3291 | GCTGCCAGTG  | CGGATAAGTC  | GTGTCTTACC  | GGTTGGACT   | CAAGACGATA  | GTTACCCGGAT | AAGCGCAGC   |
| 3361 | GGTCGGGCTG  | AACGGGGGGT  | TCGTGCACAC  | AGCCAGCTT   | GGAGCGAAACG | ACCTACACCG  | AACTGAGATA  |
| 3431 | CCTACAGCGT  | GAGCTATGAG  | AAAGCGCCAC  | GCTTCCCGAA  | GGGAGAAAGG  | CGGACACGTA  | TCCGGTAAAGC |
| 3501 | GGCAGGGTCG  | GAACAGGAGA  | GCGCACGAGG  | GAGCTTCCAG  | GGGGAACGC   | CTGGTATCTT  | TATAGTCCCTG |
| 3571 | TCGGGTTTCG  | CCACCTCTGA  | CTTGAGCGTC  | GATTTTGTG   | ATGCTCGTCA  | GGGGGCGGA   | GCCTATGGAA  |
| 3641 | AAACGCCAGC  | AACGCGGCCT  | TTTTACGGTT  | CCTGGCCCTT  | TGCTGGCCCTT | TTGCTCACAT  | GTTCTTTTCT  |
| 3711 | GCGTTATCCC  | CTGATTCTGT  | GGATAACCGT  | ATTACCGCCT  | TTGAGTGAAGC | TGATACCGCT  | CGCCGACGCC  |
| 3781 | GAACGACCGA  | GCGCAGCGAG  | TCAGTGAGCG  | AGGAAGCGGA  | AGAGCGCCCA  | ATACGCAAC   | CGCCTCTCCC  |
| 3851 | CGCGCGTTGG  | CCGATTTCATT | AATGCAGCTG  | GCACGACAGG  | TTTCCCGACT  | GGAAAGCGGG  | CAGTGAGCGC  |
| 3921 | AACGCAATTA  | ATGTGAGTTA  | GCTCACTCAT  | TAGCACCCCC  | AGGCTTTACA  | CTTTATGCTT  | CCGGCTCGTA  |
| 3991 | TGTTGTGTGG  | AATTGTGAGC  | GGATAACAAT  | TTCAACACAG  | AAACAGCTAT  | GACCATGATT  | ACGCCAAGCG  |
|      |             |             |             | KpnI        |             |             |             |
|      |             |             |             | ~~~~~       |             |             |             |
| 4061 | CGCAATTAAAC | CCTCACTAAA  | GGGAACAATA  | GCTGGGTACC  | GGGCCCCCCC  | TCGAGGTCAT  | TCATATGCTT  |
| 4131 | GAGAAGAGAG  | TCGGGATAGT  | CCAAAATAAA  | ACAAAAGTAA  | GATTACCTGG  | TCAAAAAGTGA | AAACATCAGT  |
| 4201 | TAAAAGGTGG  | TATAAGTAAA  | ATATCGGTAA  | TAAAAGGTGG  | CCCAAAGTGA  | AATTACTCT   | TTTCTACTAT  |
| 4271 | TATAAAAAAT  | GAGGATGTTT  | TGTCGGTACT  | TTGATACGTC  | ATTTTGTGAT  | GAATTGGTTT  | TTAAGTTTAT  |
| 4341 | TCGCGATTTG  | GAAATGCATA  | TCTGTATTGG  | AGTCGGTTT   | TAAGTTCTGT  | GCTTTGTGTA  | ATACAGAGGG  |
| 4411 | ATTGTATATA  | GAAATATCTT  | TAAAAAACCC  | ATATGCTAAT  | TTGACATAAT  | TTTTTGAGAAA | AATATATATT  |
|      |             | ECORI       |             |             |             |             |             |
|      |             | ~~~~~       |             |             |             |             |             |
| 4481 | CAGGCGAATT  | CCACAATGAA  | CAATAATAAG  | ATTAAAATAG  | CTTGCCCCCG  | TTGCAGCGAT  | GGGTATTTTT  |
| 4551 | TCTAGTAAAA  | TAAAAGATAA  | ACTTAGACTC  | AAAACATTTA  | CAAAAACAAC  | CCCTAAAGTC  | CTAAAGCCCCA |

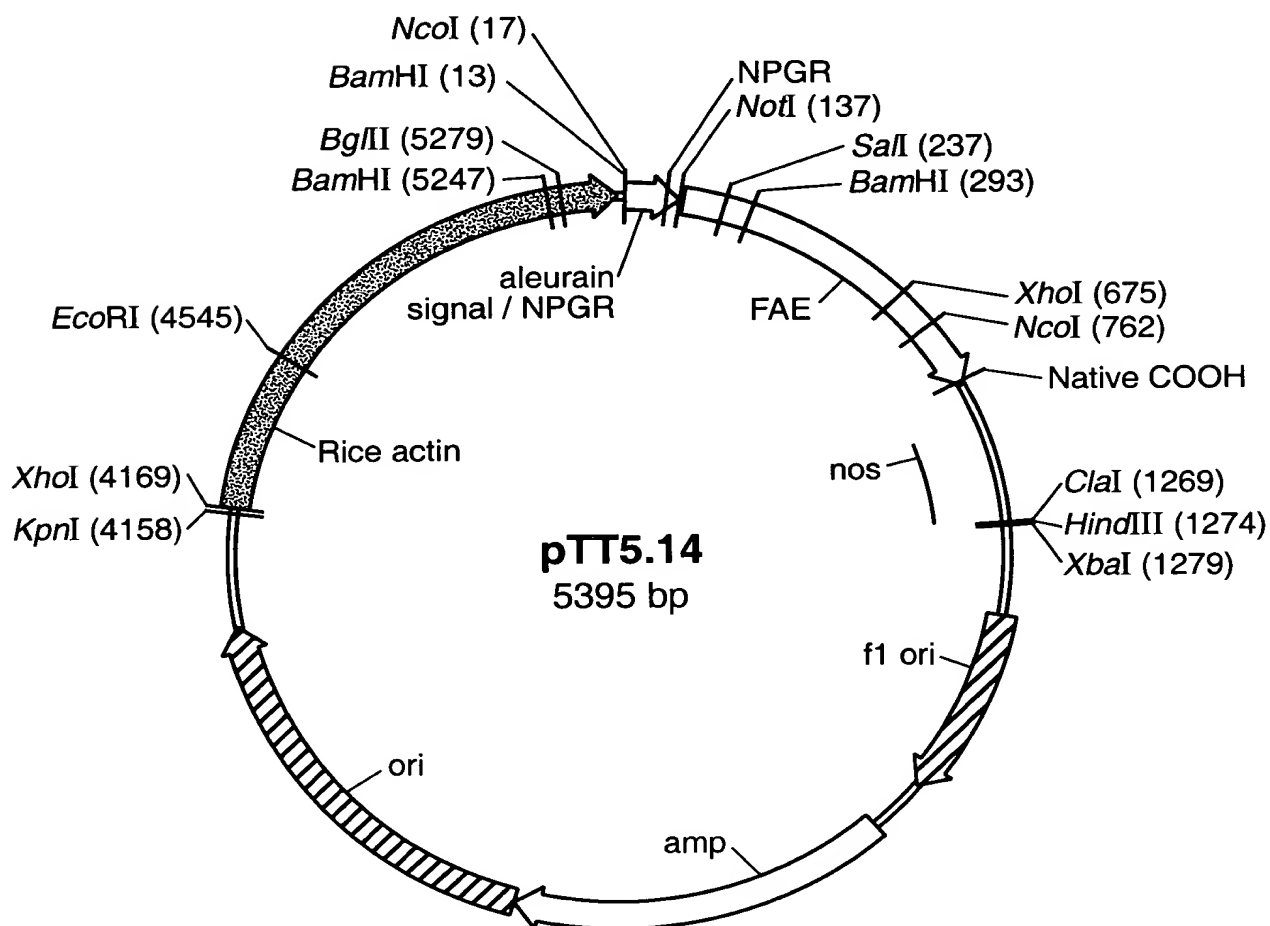
FIG.-34D

```
4621 AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCACCCCA GTGCAGCCAA
4691 CTGGCAAATA GTCTCCACCC CCGGCACATAT CACCGTGAGT TGTCCGCACC ACCGCACGTC TCGCAGCCAA
4761 AAAAAAAAAA AGAAGAAAAA AAAAGAAAAA GAAAGAACAGC AGGTGGGTCC GGTCTGTGGG GGCCGGAAAA
4831 GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC
4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCACA CCTACCACCA CCACCACCCAC CACCTCCTCC
4971 CCCCTCGCTG CCGGACGACG AGCTCCTCCC CCCTCCCCCT CCGCCGCCGC CGGTAACCCAC CCCGCCCTC
5041 TCCCTCTTCT TTCTCCGTTT TTTTCTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCCGA
5111 GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG GATCTCGCGG CTGGCGTCTC CGGGCGTGAG

 BamHI BglII
          ~~~~~          ~~~~~
5181 TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG
5251 AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG
5321 CGGAGCTTTT TTGTAGC
```

FIG.\_34E

50 / 154



**FIG.\_35A**

```

      NcoI
      ~~~~~
 BamHI
      ~~~~~
1  CCTGACGCCG AGGATCCATG GCCACGCCC GCGTCCTCCT CCTLGCGCTC GCCGTGCTGG CCACGGCCGC
   M A H A R V L L L A L A V L A T A A .
   NotI
71 CGTCGCCGTC GCCTCCTCCT CCTCCTTCGC CGACTCCAAC CCGGGCCGGC CCGTCACCGA CCGCGCGGCC
   V A V A S S S F A D S N P G R P V T D R A A
   NotI
141 GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT GGCCACTATC TCCCAAGCTG
   A S T Q G I S E D L Y S R L V E M A T I S Q A A
   SalI
211 CCTACGCCGA CCTGTGCAAC ATTCGCTCGA CTATTATCAA GGGAGAGAAA ATTACAAATT CTCAAACTGA
   Y A D L C N I P S T I I K G E K I Y N S Q T D .
   BamHI
   ~~~~~
281 CATTACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA ATAATCACCG TCTTCCGTGG CACTGGTAGT
 I N G W I L R D D S S K E I I T V F R G T G S
 D T N L Q L D T N Y T L T P F D T L P Q C N G C
351 GATACGAATC TACAACCTCGA TACTAATAC ACCCTCAGC CTTTCGACAC CCTACCACAA TGCAACGGTT
 E V H G G Y I G W V S V Q D Q V E S L V K Q .
421 GTGAAGTACA CCGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA GTCGAGTCGC TTGTCAAACA
 Q V S Q Y P D Y A L T V T G H X L G A S L A A
491 GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCCACKCC TCGGCGCCTC CCTGGCGGCA
 L T A A Q L S A T Y D N I R L Y T F G E P R S G
561 CTCACCTGCC CACAGCTGTC TCGACATAC GACAACATCC GCCTGTACAC CTTCCGGCGAA CCGCGCAGCG
 XhoI
   ~~~~~
631 GCAATCAGC CTTCGCGTCG TACATGAACG ATGCCTTCCA AGCCTCGAGC CCAGATACGA CGCAGTATTT
   N Q A F A S Y M N D A F Q A S S P D T T Q Y F .
   NotI
   ~~~~~
 R V T H A N D G I P N L P P V E Q G Y A H G G

```

FIG.\_35B

52 / 154

```

701 CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC CCGGTGGAGC AGGGGTACGC CCATGGCGGT
 V E Y W S V D P Y S A Q N T F V C T G D E V Q C
771 GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCAGAACCA CATTGTCTTG CACTGGGGAT GAAGTGCAGT
 . C E A Q G G Q G V N N A H T T Y F G M T S G A .
841 GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT TTTGGGATGA CGAGCGGAGC
 . C T W *
911 CTGTACATGG TGATCAGTCA TTTCAGCCTC CCCGAGTGTA CCAGGAAAGA TGGATGTCCT GGAGAGGGGG
981 CCGCGTAACC ACTGAAGGAT GAGCTGTAAA GAAGCAGATC GTTCAAAACAT TTGGCAATAA AGTTTCTTAA
1051 GATTGAATCC TGTGCGCGT CTGCGATGA TTATCATATA ATTTCTGTTG AATTACGTTA AGCATGTAAT
1121 AATTAAACATG TAATGCATGA CGTTATTAT GAGATGGGT TTTATGATTA GAGTCCCGCA ATTATACATT
1191 TAATACGCGA TAGAAAACAA AATATAGCGC GCAAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT
 HindIII
    ~~~~~

```

Clal XbaI

```

1261 TACTAGATCG ATAAGCTTCT AGAGCGGCGG GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTACGGC
1331 CGCTCACTGG CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT AATCGCCTTG
1401 CAGCACATCC CCTTTTCGCC AGCTGGCGTA ATAGCGAAGA GCGCCGCACC GATCGCCCTT CCCAACAGTT
1471 GCGCAGCCTG AATGGCGAAT GGGACGCGCC CTGTAGCGGC GCATTAAGCG CGCGGGGTGT GGTGGTTACG
1541 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCC CTCTTTTCGC TTTCTTCCCT TCCTTTCTCG
1611 CCACGTTTCG CCGCTTTCCC CGTCAAGCTC TAAATCGGG GCTCCCTTTA GGTTCCGAT TTAGTGCTTT
1681 ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGATAGT GGCATCGCC CTGATAGACG
1751 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAACT GGAACAACAC
1821 TCAACCCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAAA
1891 TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAACAAA ATATTAAACG TTAACAATTA GTGGCACATT
1961 TTCGGGAAA TGTGCGCGGA ACCCTATT TTTTATTTT CTAATAACAT TCAAAATATG ATCCGCTCAT
2031 GAGACAATAA CCTGATAAA TGCTTCAATA ATATTGAAA AGGAAGAGTA TGAGTATTCA ACATTTCCGT
2101 GTCGCCCTTA TTCCCTTTT TCGCGCATTT TGCCTTCCCTG TTTTGTCTCA CCCAGAAACG CTGGTGAAG
2171 TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAACTG GATCTCAACA GCGGTAAGAT
2241 CCTTGAGAGT TTTTCGCCCG AAGAACGTTT TCCAATGATG AGCATTTTA AAGTTCTGCT ATGTGGCGCG
2311 GTATTATCCC GTATTGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA CTATTCTCAG AATGACTTGG
2381 TTGAGTACTC ACCAGTCACA GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAAATTAT GCAGTGCTGC
2451 CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAAC
2521 GCTTTTTTGC ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA
2591 TACCAAAACGA CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAACG TTGCGCAAC TATTAAGTGG
2661 CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA

```

FIG.\_35C



53 / 154

```

2731 CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC
2801 GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG
2871 TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGGTAA
2941 CTGTCAGACC AAGTTTACTC ATATATACTT TAGATTGATT TAAAACTTCA TTTTAAATTT AAAAGGATCT
3011 AGGTGAAGAT CCTTTTGTAT AATCTCATGA CCAAAATCCC TTAACGTGAG TTTTCGTTCC ACTGAGCGTC
3081 AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT TTGAGATCCT ATCAAGAGCT ACCAACTCTT TTTCCGAAGG
3151 ACAAATAAAC CACCGCTACC AGCGGTGGTT TGTGTGCGG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG
3221 TAAGTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCTT CGCTCTGCTA ATCCGTGTAC CAGTGGCTGC TGCCAGTGGC
3291 CAAGAACTCT GTAGCACCGC GTTGGACTCA AGCAGCTTGG AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA
3361 GATAAGTCGT GTCTTACCGG GTCGACTTGG GAGAAAGCG GGTATCTTTA TAGTCCCTGC GGTATCCCTT
3431 CGGGGGTTC GTGCACACAG CCCAGCTTGG GTTCCGAGG GCTTCCAGG GGTATCTTTA TAGTCCCTGC GGTATCCCTT
3501 GCTATGAGAA AGCGCCACGC GCACGAGGGA GCTTCCAGG TTTTGTGTAT GCTCGTCCG GGTATCTTTA TAGTCCCTGC
3571 ACAGGAGAGC GCACGAGGGA GCTTCCAGG TTTTGTGTAT GCTCGTCCG GGTATCTTTA TAGTCCCTGC
3641 ACCTCTGACT TGAGCGTCCA TTTTGTGTAT GCTCGTCCG GGTATCTTTA TAGTCCCTGC GGTATCCCTT
3711 CGCGGCTTTT TTACGGTTCC TGCGCTTTTG TGGCCCTTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC
3781 GATTCGTGG ATAAACCGTAT TACCGCTTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC
3851 GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAAACG CCTCTCCCCG CGCGTTGGCC
3921 GATTCATTAA TGCAGCTGGC ACGACAGGTT TCCCGACTGG AAAGCGGCA GTGAGCGCAA GCCAATTAAT
3991 GTGAGTTAGC TCACTCATTA GGCACCCCG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
4061 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCGCG CAATTAACCC

                                KpnI                                XhoI
~~~~~                                ~~~~~
4131 TCACTAAAGG GAACAAAAGC TGGGTACCGG GCGCCCTCTC GAGGTCAATC ATATGCTTGA GAAGAGAGTC
4201 GGGATAGTCC AAAATAAAAC AAAGGTAAGA TTACCTGGTC AAAAGTGAAA ACATCAGTTA AAAGGTGGTA
4271 TAAGTAAAT ATCGGTAATA AAAGGTGGCC CAAAGTGAAA TTTACTCTTT TCTACTATTA TAAAAAATGA
4341 GGATGTTTTG TCGGTACTTT GATACGTCAT TTTTGTATGA ATTGGTTTTT AAGTTTATTC GCGATTGGGA
4411 AATGCATATC TGTATTTGAG TCGGTTTTTA AGTTCGTTGC TTTTGTAAAT ACAGAGGGAT TTGTATAAGA

 EcoRI
~~~~~
4481 AATATCTTTA AAAAACCCAT ATGCTAAATT GACATAAATT TTGAGAAAAA TATATATTCA GGCGAATTCC
4551 ACAATGAACA ATAATAAGAT TAAAATAGCT TGCCCCCGTT GCAGCGATGG GTATTTTTTC TAGTAAAAA
4621 AAAGATAAAC TTAGACTCAA AACATTTACA AAAACAACCC CTAAGTCCCT AAAGCCCAAA GTGCTATGCA
4691 CGATCCATAG CAAGCCCGAG CCAACCCCAAC CCAACCCCACT GCAGCCCACT GGCAATAATAGT
4761 CTCACCCCCC GGCATATCA CCGTGAGTTG TCCGCACCCAC CGCAGCTCTC GCAGCCCAAA AAAAAAAG
4831 AAAGAAAAAA AAGAAAAAGA AAAACAGCAG GTGGGTCCCG GTCTGTGGGG CCGGAAAAAGC GAGGAGGATC

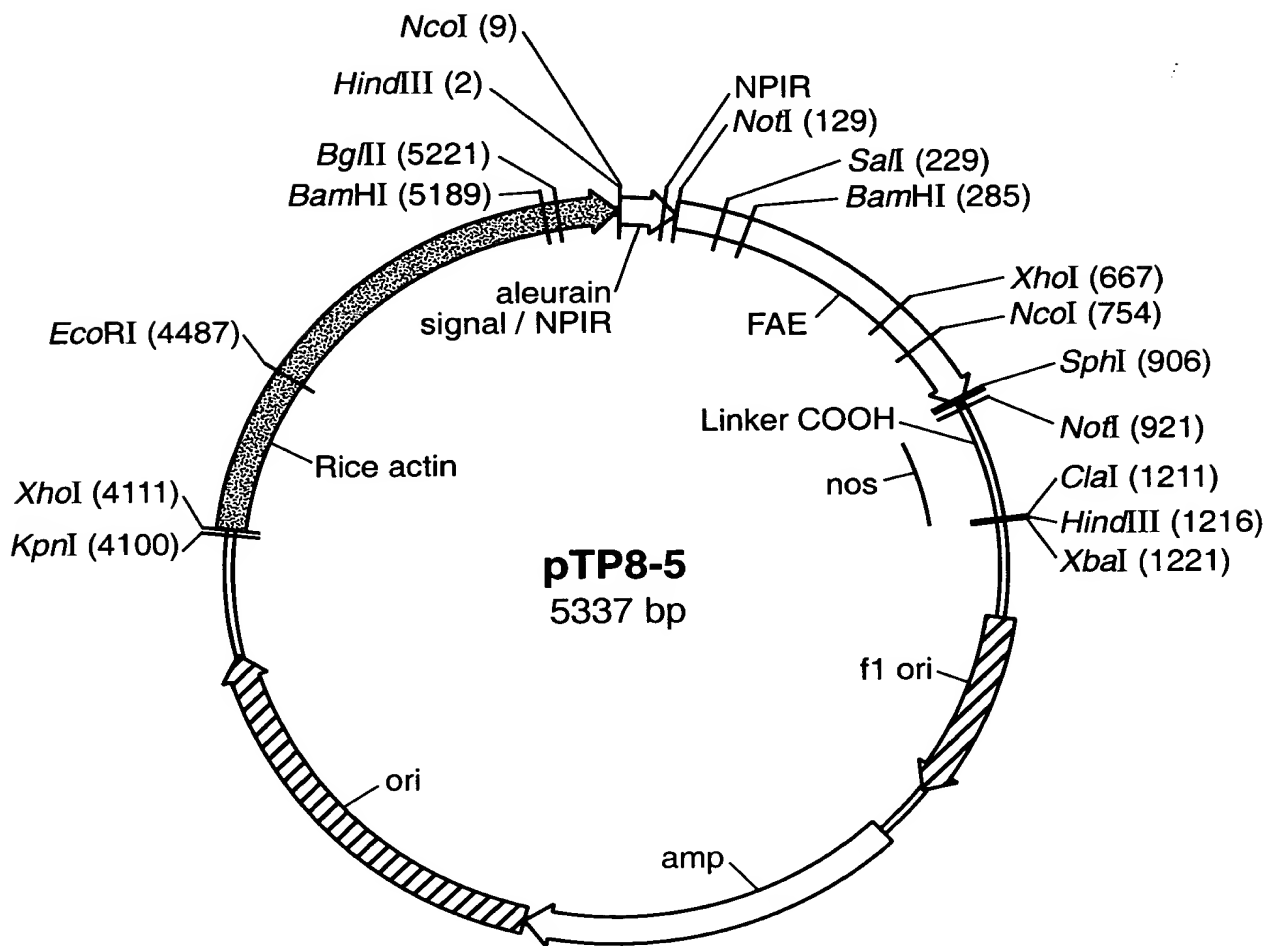
```

FIG.\_35D

4901 GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC CGCTTCCAA GAAACGCCCC CCATCGCCAC TATATACATA  
4971 CCCCCCCTC TCCTCCCATC CCCCCAACC TACCACCACC ACCACCACCA CCTCCTCCCC CCTCGCTGCC  
5041 GGACGACGAG CTCTCCCTCC CTCCCTCCCTC GCGCGCGCG GTAAACCACCC CGCCCCCTCTC CTCCTTCTTT  
5111 CTCCGTTTTT TTTTTCGTCT CGGTCTCGAT CTTTGGCCTT GGTAGTTGG GTGGGCGAGA GCGGCTTCGT  
BamHI  
5181 CGCCACAGATC GGTGCGCGGG AGGGCGGGA TCTCGCGGCT GCGTCTCCG GCGGTGAGTC GGCCCGGATC  
BamHI  
~  
5251 CTCGCGGGA ATGGGCTCT CGGATGTAGA TCTTCTTCT TTCTTCTTTT TGTGGTAGAA TTTGAATCCC  
5321 TCAGCATTGT TCATCGGTAG TTTTCTTTT CATGATTGT GACAAATGCA GCCTCGTGCG GAGCTTTTTT  
5391 GTAGC

FIG.\_35E

55 / 154



**FIG.\_36A**

2  
2  
2  
2  
2  
2

2  
2  
2  
2  
2  
2

**Not I**

# Salt

2  
2  
2  
2  
2  
2

## BamHI

2  
2  
2  
2  
2

CONFIDENTIAL

W A C Y V T C W Y S Y O D O V F S I V K O O V S

. O Y P D Y A L T V T G H X L G A S L A A L T A .

CONFIDENTIAL

**Index**

五五五五五

**NcoI**

2  
2  
2  
2  
2  
2

. H A N D G I P N L P P V E Q G Y A H G G V E Y .

57 / 154

```

701 CTCATGCCAA CGACGGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA
    . W S V D P Y S A Q N T F V C T G D E V Q C C E
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACCTGGG ATGAAGTGCA GTGCTGTGAG
    SphI
    ~~~~
 A Q G G Q G V N N A H T T Y F G M T S G A C T W
841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT
 NotI
    ~~~~~~
    . P V A A A *
911 GGCCGGTCGC GGCCGCGTAA CCACTGAAGG ATGAGCTGTA AAGAAAGCAGA TCGTTCAAAC ATTTGGCAAT
981 AAAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGCGAT GATTATCATA TAAATTCGTGT TGAATTACGT
1051 TAAGCATGTA ATAATTAAACA TGTAAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG
1121 CAATTATACA TTTAATACGC GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG
    HindIII
    ~~~~~~
 ClaI XbaI
    ~~~~~~
1191 TGTCACTCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC CTATAGTGAG
1261 TCGTATTACG CGGCTCACT GGCGTCTGTT TTACAACGTC GTGACTGGGA AAACCTTGGC GTTACCCCAAC
1331 TTAATCGCCT TGCAGCACAT CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG GCGCATTAAG CCGCGCGGGT
1471 GTGGTGGTTA CGCGCAGCGT GACCCTTACA CTTGCCAGCG CCTAGCGCC CGCTCCTTTC GCTTTCCTCC
1541 CTTCCCTTCT CGCCACGTTT TCGACCCCAA AAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG
1611 ATTTAGTGCT TTACGGCACC CCCTTTTCG ATCTCGGTCT TTGGAGTCCA CGTTCCTTAA TAGTCCCAA
1681 CCCTGATAGA CGGTTTTCG CCCTTTTCG ATCTCGGTCT TTGGAGTCCA CGTTCCTTAA TAGTCCCAA
1751 CTGGAACAAC ACTCAACCTT AATGAGCTGA TTTAACAAA AATTAACGCG AATTTTAAAC AAATATAAT
1821 TTGGTTAAA AATGAGCTGA TTTAACAAA AATTAACGCG AATTTTAAAC AAATATAAT
1891 TAGGTGGCAC TTTTCGGGA AATGAGCTGA TTTAACAAA AATTTTAAAC AAATATAAT
1961 GTATCCGCTC ATGAGACAAT AACCTTGATA AATGCTTCAA AATGCTTCAA AATGCTTCAA AATGCTTCAA
2031 CAACATTTCC GTGTCGCCCT TATTCCCTTT TATTCCCTTT TATTCCCTTT TATTCCCTTT TATTCCCTTT
2101 CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC TTTGCCCTTC TTTGCCCTTC TTTGCCCTTC
2171 CAGCGGTAG ATCCTTGAGA GTTTTCGCC CCAAGAACGT TTTCCCAATGA TGAGCACTTT TAAAGTTCTG
2241 CTATGTGGCG CCGTATTATC CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC
2311 AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCAATGACAG TAAGAGAAAT

```

FIG.\_36C

58 / 154

```

2381 ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAAACGAT CGGAGGACCG
2451 AAGGAGCTAA CCGCTTTTTC GCACAAACATG GGGGATCATG TAACTCGCCT TGAATCGTTGG GAACCGGAGC
2521 TGAATGAAGC CATACCAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA
2591 ACTATTAACT GCGAAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGA GCGGATATAA
2661 GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTAATTGC TGATAAATCT GGAGCCGGTG
2731 AGCGTGGGTC TCGCGGTATC ATTGCAGCAC ATATGGATGA ACGAAATAGA TGGTAAGCCC TCCGTAATCG TAGTTATCTA
2801 CACGACGGGG AGTCAGGCAA AACTGTGAGA CCAAGTTTAC TCATATATAC CAGATCGCTG AGATAGTGC CTCACCTGATT
2871 AAGCATTTGGT AACTGTGAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACTT CATTTTAAAT
2941 TTAAGAGGAT CTAGGTGAAG ATCCTTTTGG TAGAAAAGAT CAAAGGATCT TCTTGAGATC GAACAAAATC CTTTTCGTT
3011 CCACTGAGCG TCAGACCCCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC GAACAAAATC CTTTTCGTT
3081 TGCTGCTTGC AAACAAAATA GGTAACTGGC TTCAAGCAAG ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC
3151 TTTTCCGAA AGGCCACCA TTTCAAGAACT CTGTAGCAC TTCAAGCAAG ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC
3221 AGGCCACCA TTTCAAGAACT CTGTAGCAC TTCAAGCAAG ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC
3291 GCTGCCAGTG GCGATAAGTC AACGGGGGT TCCTGCTTACC GGTGAGTGC GCCTACATAC CTCGCTCTGC TAATCCCTGTT ACCAGTGGCT
3361 GGTCCGGCTG AACGGGGGT TCCTGCTTACC GGTGAGTGC GCCTACATAC CTCGCTCTGC TAATCCCTGTT ACCAGTGGCT
3431 CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC
3501 GGCAGGGTCG GAACAGGAGA GCGACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCCTG
3571 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG ATGCTCGTCA GGGGGCGGA GGTCTTCCCT
3641 AAACGCCAGC AACCGGCCCT TTTTACGGTT CTTGAGCGTC GATTTTGTG ATGCTCGTCA GGGGGCGGA GGTCTTCCCT
3711 GCCTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCTT TTTGAGTGC TTTGAGTGC TTTGAGTGC TTTGAGTGC
3781 GAACGACCGA GCGACGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAC CGCTCTCTCC
3851 CGCGCGTTGG CCGATTCAAT AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC
3921 AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA
3991 TGTTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG AACACAGCTAT GACCATGATT ACGCCAAGCG

4061 CGCAATTAA C CTTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCAGAGTCAT TCATATGCTT
4131 GAGAAGAGAG TCGGGATAGT CCATAATAA ACATAAGTAA GATTACCTGG TCAAAAAGTGA AAACATCAGT
4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCATAAGTGA AATTACTCT TTTCTACTAT
4271 TATAAAAAAT GAGGATGTTT TGTCGGTACT TTGATACGTC ATTTTTGTAT GAATTGGTTT TTAAGTTTAT
4341 TCGCGATTGG GAAATGCATA TCTGTATTG AGTCGGTTT TAAAGTTCGTT GCTTTTGTAA ATACAGAGGG
4411 ATTTGTATTA GAAATATCTT TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATATT

      EcoRI
      ~~~~~
4481 CAGGCGAATT CCACAATGAA CAATAATAAG ATTAAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT

```

XhoI

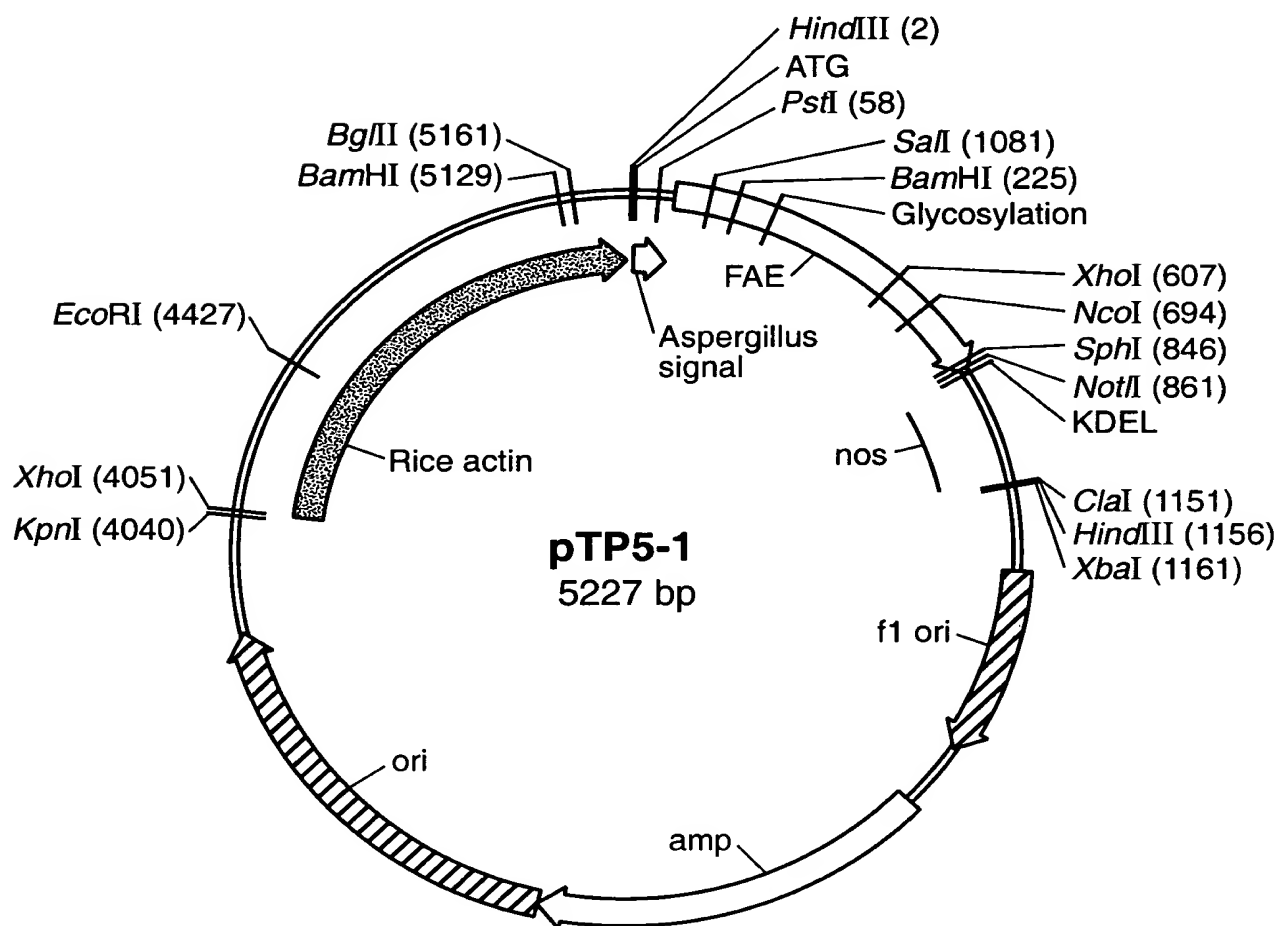
KpnI

FIG.\_36D

```
4551 TCTAGTAAAA TAAAAGATAA ACTTAGACTC AAAACATTTA CAAAAACAAC CCCTAAAGTC CTAAAGCCCA
4621 AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCAACCCA GTGCAGCCAA
4691 CTGGCAAATA GTCTCCACCC CCGGCACTAT CACCGTGAGT TGTCGGCACC ACCGCACGTC TCGCAGCCAA
4761 AAAAAAATAA AGAAAGAAAA AAAGAAAAA AAAGAAAAA GAGTGGGTCC GGGTCGTGGG GGCCGGAAAA
4831 GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC
4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCAAC CTAACCAACA CCACCAACAC CACCTCCTCC
4971 CCCCTCGCTG CCGGACGACG AGTCCTCCCC CCCTCCCCCT CCGCCGCCGC CGGTAACCAAC CCCGCCCTC
5041 TCCTCTTTCT TTCTCCGTTT TTTTTCGTCG CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCGA
5111 GAGCGGCTTC GTCGCCCA GA TCGGTGCGCG GGAGGGGCGG GATCTCGCGG CTGGCGTCTC CGGGCGTGAG
 BamHI
                                     ~~~~~
5181 TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTCCTTCTT TTTGTGGTAG
5251 AATTGAATC CCTCAGCAAT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG
5321 CGGAGCTTTT TTGTAGC
```

FIG. 36E

60 / 154



**FIG.\_37A**



61 / 154

```
HindIII
~~~~~
 M K Q F S A K H V L A V V V T A G H A L A
1 AAGCTTAACA TGAAGCAGTT CTCCGCCAAA CAGTCTCTCG CAGTTGTGGT GACTGCAGGG CACGCCCTTAG
 . A S T Q G I S E D L Y S R L V E M A T I S Q A .
71 CAGCCTCTAC GCAAGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC

 Sali
      ~~~~~
      . A Y A D L C N I P S T I I K G E K I Y N S Q T
141 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAACT

      BamHI
      ~~~~~
 D I N G W I L R D D S S K E I I T V F R G T G S
211 GACATTAACG GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA
 . D T N L Q L D T N Y T L T P F D T L P Q C N G .
281 GTGATACGAA TCTACAACATC GATACTAACT ACACCTCTAC GCCTTTCGAC ACCTACCAC AATGCAACGG
 . C E V H G G Y I G W V S V Q D Q V E S L V K
351 TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA
 Q Q V S Q Y P D Y A L T V T G H X L G A S L A A
421 CAGCAGGTTA GCCAGTATCC GGAATACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG
 . L T A A Q L S A T Y D N I R L Y T F G E P R S .
491 CACTCACTGC CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAT ACCTTCGGCG AACCGCGCAG

 XhoI
      ~~~~~
      . G N Q A F A S Y M N D A F Q A S S P D T T Q Y
561 CGGCAATCAG GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCCTCGA GCCAGATAC GACGCAGTAT

      NcoI
      ~~~~~
 F R V T H A N D G I P N L P P V E Q G Y A H G G
631 TTCCGGGTCA CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCCTGGA GCAGGGGTAC GCCATGGCG
 . V E Y W S V D P Y S A Q N T F V C T G D E V Q .
701 GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTCT TGCACCTGGG ATGAAGTGCA
 . C C E A Q G G Q G V N N A H T T Y F G M T S G
771 GTGCTGTGAG GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC
```

FIG.\_37B

**FIG.\_37C**

63 / 154

```

2731 TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC
2801 CTCACATGATT AAGCATTTGGT AACTGTGCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACTT
2871 CATTTTAAAT TTAATAAGGAT TTAATAAGGAT ATCCTTTTGG TAGAAAGAT CAAAGGATCT TCTTGAGATC CTTTATTCT
2941 AGTTTTCGTT CCACTGAGCG TGCTGCTTGC AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3011 GCGGTAAATC GTCTGCTTGC TTTTTCGGAA GGTAACTGGC TTCAAGAACT CTGTAGACAC GCTTACCTGTT TAATCCTGTT
3081 CTACCAATCT AGCCGTAAGT AGCCACCCAC GTGCCAGTG GCGATAAGTC CTTGACGAGT CAAGACGATA CTTACCCGGAT
3151 AGCCGTAGTT AGCCGTAAGT AGCCACCCAC GTGCCAGTG GCGATAAGTC CTTGACGAGT CAAGACGATA CTTACCCGGAT
3221 ACCAGTGGCT GCTGCCAGTG GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3291 AAGGCGCAGC GGTGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3361 AACTGAGATA CCACTGAGCGT GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3431 TCCGGTAAAG GGCAGGCTG GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3501 TATAGTCCCTG TCGGCTTTCG GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3571 GCCTATGGAA AACGCCAGC AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3641 GTTCTTTTCTT GCGTTATCCC GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3711 CGCCGACGCC GAACGACCGA GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT CTTGTTTGGC GGATCAAGAG
3781 CGCCTCTCCC GCGCGCTTGG CCGATTTCAT AATGACGCTG GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT
3851 CAGTGAGCGC AACGCAATTA ATGTGAGTTA GTCGCGGCTG GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT
3921 CCGGCTCGTA TGTGTGTGTG AATTGTGAGC GTCGCGGCTG GTCGCGGCTG AACTGCTTGC ACCACCGCTA CCAAGGATCT

3991 ACGCCAAAGC CGCAATTAAAC CCTCACTAAA GGAACAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTGAT
4061 TCATATGCTT GAGAAAGAGAG TCGGGATAGT CCATAAATAA ACAAAAGGTAA GATTACCTGG TCAAAAAGTGA
4131 AAACATCAGT TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCAAAAGTGA AATTACTCT
4201 TTTCTACTAT TATAAAAATT GAGGATGTTT TGTCGGTACT TTGATACGTC ATTTTTGTAT GAATTGGTGT
4271 TTAAGTTTAT TCGCGATTGG GAAATGCATA TCTGTATTGG AGTCGGTTTT TAAGTTCGTT TTTTGTGATA
4341 ATACAGAGGG ATTTGTATAA GAAATATCTT TAAAAAACC ATATGCTAAT TTGACATAAT TTTTGAGAAA

ECORI
4411 AATATATATT CAGGCGAATT CCACAATGAA CAATAATAAG ATTAATAATAG CTTGCCCCCG TTGCAGCGAT
4481 GGGTATTTT TCTAGTAAA TAAAAGATAA ACTTAGACTC AAAACATTTA CAAAACAAC CCCTAAAGTC
4551 CTAAGGCCCA AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCAACCCA
4621 GTGCAGCCCA CTGGCAATA GTCTCCACC CCGGCACTAT CACCGTGAGT TGTCCGCACC ACCGCACGTC
4691 TCGCAGCCCA AAAAAAATA AGAAAGAAA AAAAAACAGC AGGTGGGTCC AGGTGGGTCC AGGTGGGTCC

```

FIG. 37D

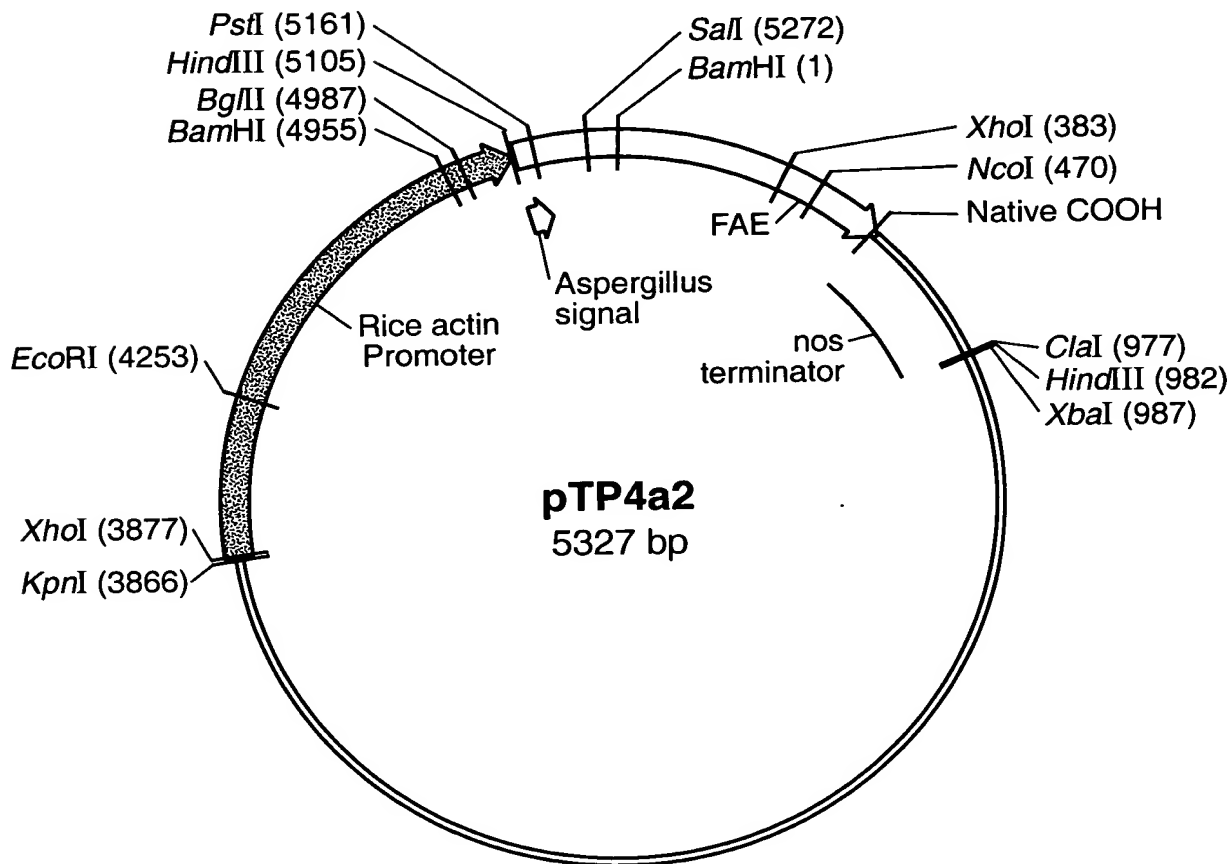
09/09/2009 07:02:02

64 / 154

```
4761 GCGCGGAAA GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC
4831 CCCCATCGCC ACTATATACA TACCCCCCCC TCCTCTCCCA TCCCCCAAC CTTACCACCA CCACCACCAC
4901 CACCTCCTCC CCCCTCGCTG CCGGACGACG AGCTCCTCCC CCCTCCCCCT CGCGCGCCGC CGGTAACACAC
4971 CCGCCCCCTC TCCTCTTCT TCTCCTGTTT TTTTCTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT
5041 GGGTGGGCGA GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG GATCTCGCGG CTGGCGTCTC
 BamHI
      ~~~~~
5111 CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT
5181 TTTGTGGTAG AATTGAATC CCTCAGCAAT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG
5251 CAGCCTCGTG CGGAGCTTTT TTGTAGC
      BglII
      ~~~~~
```

FIG.\_37E

65 / 154



**FIG.\_38A**

66 / 154

```
BamHI
~~~~~
. I L R D D S S K E I I T V F R G T G S D T N L
1  GATCCTCCGC GACGACAGCA GCAAAGAAAT AATCACCCTC TTCCGTGGCA CTGGTAGTGA TACGAATCTA
  Q L D T N Y T L T P F D T L P Q C N G C E V H G
71  CAACTCGATA CTAACACTAC CCTCAGCCCT TTCGACACCC TACCACAATG CAACGGTTGT GAAGTACACG
  . G Y Y I G W V S V Q D Q V E S L V K Q Q V S Q .
141 GTGGATATTA TATTGGATGG GTCTCCGTCC AGGACCAAGT CGAGTCGCTT GTCAAAACAGC AGGTTAGCCA
  . Y P D Y A L T V T G H X L G A S L A A L T A A
211 GTATCCGGAC TACGGCTGA CCGTGACCGG CCACKCCCTC GCGCCCTCCC TGGCGGCACT CACTGCCGCC
  Q L S A T Y D N I R L Y T F G E P R S G N Q A F
281 CAGCTGTCTG CGACATACGA CAACATCCGC CTGTACACCT TCGGCGAACC GCGCAGCGGC AATCAGGCCT
      xhoI
      ~~~~~
. A S Y M N D A F Q A S S P D T T Q Y F R V T H .
351 TCGCGTCGTA CATGAACGAT GCCTTCCAAG CCTCGAGCCC AGATACGACG CAGTATTTC GGTCACTCA
 NcoI
      ~~~~~
. A N D G I P N L P P V E Q G Y A H G G V E Y W
421 TGCCAACGAC GGCATCCCA ACCTGCCCCC GGTGAGCAG GGTACGCCC ATGGCGGTGT AGAGTACTGG
  S V D P Y S A Q N T F V C T G D E V Q C C E A Q
491 AGCGTTGATC CTTACAGCGC CCAGAACACA TTGTCTGCA CTGGGGATGA AGTGCAGTGC TGTGAGGCC
  . G G Q G V N N A H T T Y F G M T S G A C T W * .
561 AGGCGGACA GGGTGTGAAT AATGCGCACA CGACTTATTT TGGGATGACG AGCGGAGCCT GTACATGGTG
  . *
631 ATCAGTCATT TCAGCCTCCC CGAGTGTACC AGGAAAGATG GATGTCTGG AGAGGGGCC GCGTAACCAC
701 TGAAGGATGA GCTGTAAAGA AGCAGATCGT TCAAACATTT GGCAATAAAG TTTCTTAAGA TTGAATCCCTG
771 TTGCCGGTCT TGCATGATT ATCATATAAT TTCTGTTGAA TTACGTTAAG CATGTAATAA TTAACATGTA
841 ATGCATGACG TTATTATATGA GATGGGTTTT TATGATTAGA GTCCCGCAAT TATACATTTA ATACGCGATA
      ClaI
911 GAAAACAAA TATAGCGCGC AAACATAGGAT AAATTATCGC GCGCGGTGTC ATCTATGTTA CTAGATCGAT
      XbaI
      ~~~~~
HindIII
~~~~~
```

FIG.-38B

981 AAGCTTCTAG AGCGGCCGGT GGAGCTCCAA TTCGCCCTAT AGTGAGTCGT ATTACGGCGG CTCACCTGGCC  
 1051 GTCGTTTTAC AACGTCGTGA CTGGGAAAC CCTGGCGTTA CCCAACTTAA TCGCCCTTGCA GCACATCCCC  
 1121 CTTTCGCCAG CTGGCGTAAT AGCGAAGAG CCCGCACCGA TCGCCCTTCC CAACAGTTGC CACAGCTGAA  
 1191 TGGCGAATGG CACGCGCCCT GACGCGCGC ATTAAGCGG CCGGGGTGCG GCGGTACGCG CAGCGTGACC  
 1261 GCTACACTTG CCAGCGCCCT CACGCGCGC ATTAAGCGG CCGGGGTGCG GCGGTACGCG CAGCGTGACC  
 1331 GCTTCCCCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG GTTCCGATTT AGTGCTTTAC GGCACCTCGA  
 1401 CCCCCAAAA CTTGATTAGG AGTCCACGTT CTTTAATAGT GGACTCTTGT TCCAAACTGG AACCTATCT  
 1471 TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT TCCAAACTGG AACCTATCT  
 1541 CGGTCATATC TTTTGATTTA TAAGGATTTT TGCCGATTTC GGCCTATTGG TTAATAAATG AGCTGATTTA  
 1611 ACAAAAATTT AACGCGAATT TTAACAAAAT TTAACGCTT ATTAACGCTT AAATATGAT CCGGGAATG  
 1681 TGCGCGGAAC CCTATTTGT TTAATTTTCT AAATACATTC AAATATGAT CCGGGAATG  
 1751 CTGATAAATG CTTCAATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAC ATTTCCGTGT CGCCCTTATT  
 1821 CCGTATTTTG CGGCAATTTG CTTCCCTGTT CTTCCCTGTT TTTGCTCACC TCGAAGTGA TCTCAACAGC GTTGAAGTTT  
 1891 AAGATCAGTT GGGTGACAGA GAACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGGCGCGGT ATTATCCCGT  
 1961 TCGCCCCGAA AAGATCAGTT GGGTGACAGA GAACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGGCGCGGT ATTATCCCGT  
 2031 ATTGACGCGG AAGATCAGTT GGGTGACAGA GAACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGGCGCGGT ATTATCCCGT  
 2101 CAGTCACAGA AAGATCAGTT GGGTGACAGA GAACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGGCGCGGT ATTATCCCGT  
 2171 TGATAACACT GCGGCCAACT ATCATGTAAC TCGCCCTGAT TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG  
 2241 AACATGGGGG ATCATGTAAC TCGCCCTGAT TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG  
 2311 AGCGTGACAC CACGATGCTT CCGCAACAAAT TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG  
 2381 TCTAGCTTCC CCGCAACAAAT TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG TAATAGACTG  
 2451 GCGCTTCCCG CTGGCTGGTT TATGCTGAT AAGCCCTCCC AAGCCCTCCC AAGCCCTCCC AAGCCCTCCC AAGCCCTCCC  
 2521 CAGCACTGGG GCCAGATGGT GCCAGATGGT GCCAGATGGT GCCAGATGGT GCCAGATGGT GCCAGATGGT GCCAGATGGT  
 2591 GGATGAACGA AATAGACAGA AATAGACAGA AATAGACAGA AATAGACAGA AATAGACAGA AATAGACAGA AATAGACAGA  
 2661 GTTACTCAT ATATACTTAA TCTCATGACC AAAATCCCTT GATGATTTA AATGATTTA AATGATTTA AATGATTTA  
 2731 TTTTGTATAA GGATCTTCTT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT  
 2801 AAAGATCAAA GGATCTTCTT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT  
 2871 CCGCTACCCG CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT CCGTGCTTCT  
 2941 GCAGAGCGCA GATACCAAAT ACTGCTCTTC TAGTGATGCC GTAGTTAGGC CACCACTTCA AGAATCTGT  
 3011 AGCACCGCCT ACATACCTCG CTCTGCTAAT CCGGATAAGG CGCAGCGGT CAGCGGTGAG GGGGGTTCGT  
 3081 CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG CGCAGCGGT CAGCGGTGAG GGGGGTTCGT  
 3151 GCACACAGCC CAGCTTGGAG CGAAGGACCT ACACCGAAT GAGATACCTA GAGATACCTA GAGATACCTA GAGATACCTA  
 3221 CGCCACGCTT CCGGAAAGGA GAAAGGCGGA CAGGTATCCG GTAAAGCGGA GGTTCGGAAC AGGAGAGCGC  
 3291 ACGAGGGAGC TTCCAGGGGG AAACGCTGG TATCTTTATA GTCTGTGCG GTTTCGCGAC CTCTGACTTG  
 3361 AGCGTCGATT TTTGTGATGC TCGTCAGGGG GCGGAGCCT ATGGAATAAC GCCAGCAACG CGGCCCTTTT

FIG.-38C

+

**FIG. 38D**



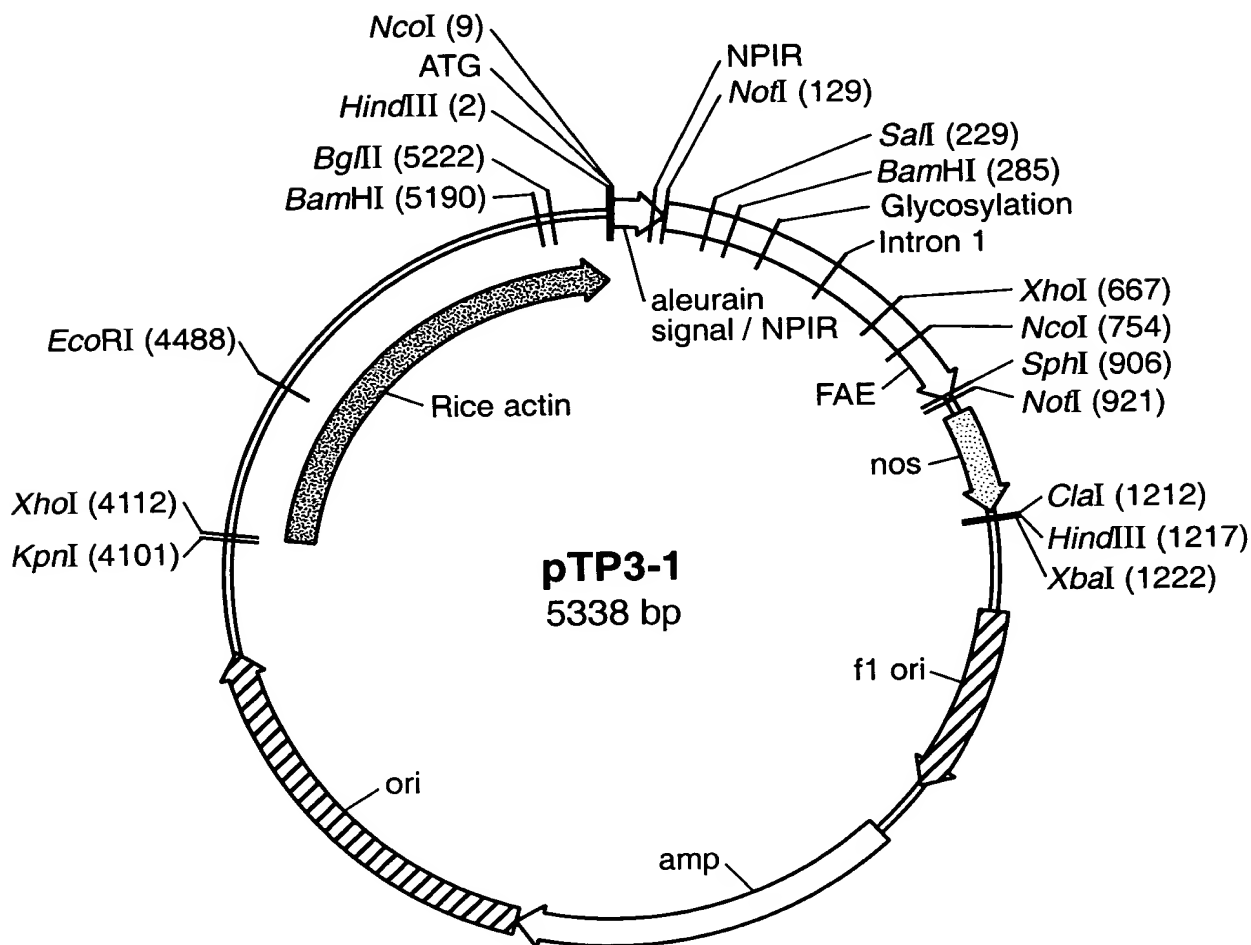
69 / 154

```

                    PstI
                    ~~~~~
5111 M K Q F S A K H V L A V V T A G H A L A A S .
 ACATGAAGCA GTTCTCCGCC AAACACGTCC TCGCAGTTGT GGTGACTGCA GGGCAGCCCT TAGCAGCCTC
 . T Q G I S E D L Y S R L V E M A T I S Q A A Y
5181 TACGCAAGGC ATCTCCGAAG ACCTCTACAG CCGTTAGTC GAAATGGCCA CTATCTCCCA AGCTGCCCTAC
 Sali
      ~~~~~
5251  A D L C N I P S T I I K G E K I Y N S Q T D I N
      GCCGACCTGT GCAACATTCC GTCGACTATT ATCAAGGGAG AGAAAATTTA CAATTCTCAA ACTGACATTA
      B
      ~
5321  . G W
      ACGGATG
```

FIG.\_38E

70 / 154



**FIG.\_39A**

71 / 154

```

NcoI
~~~~~
HindIII
~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCCGTCGCCG

      . A S S S F A D S N P I R P V T D R A A A S T .
71 TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCGATCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC
      . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC

      NPIR
      ~~~~~
 NotI
      ~~~~~

      D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAAT GACATTAACG

      BamHI
      ~~~~~

 . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTCCGT GGCACCTGGTA GTGATACGAA

 Glycosylation
      ~~~~~

      . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAACTC GATACTAACT ACACCCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
      H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA
      . Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
      . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG

```

FIG.\_39B

72 / 154

```

XhoI
~~~~~
631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T
 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGGTCA

NcoI
~~~~~
701  . H A N D G I P N L P P V E Q G Y A H G G V E Y .
      CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCCTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA
      . W S V D P Y S A Q N T F V C T G D E V Q C C E
      771  CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACCTGGG ATGAAGTGCA GTGCTGTGAG

SphI
~~~~~
841 A Q G G Q G V N N A H T T Y F G M T S G A C T W
 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT

NotI
~~~~~
KDEL
~~~~~
911 . P V A A A E T T E G *
 GGCCGGTCGC GGCCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTGGCAA
981 TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGCGA TGATTATCAT ATAAATTTCTG TTGAATTACG
1051 TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC
1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAAACT AGGATAAATT ATCGCGCGCG

HindIII
~~~~~
ClaI
~~~~~
XbaI
~~~~~
1191 GTGTCATCTA TGTTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTCG CCTATAGTGA
1261 GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTTACAACGT CGTGACTGGG AAAACCCCTGG CGTTACCCAA
1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC
1401 CTTCCCAACA GTTGCGCAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTA GCGCGCGCGG
```

FIG.-39C

73 / 154

1471 TGTGGTGGTT ACGCGCAGCG TGACCGGTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTCTTTC  
1541 CCTTCCCTTC TCGCCACGTT TCGCCGGCTT CCGCGTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
1611 GATTAGTGC TTACGGCAC CTCGACCCCA AAAAAGCTGA TTAGGGTGAT GGTTCACGTA GTGGCCATC  
1681 GCCCTGATAG ACGGTTTTC CCGCTTTCG TATCTCGGTC TATCTTTTG ATTTATAAGG GATTTGCGG ATTTCCGGCT  
1751 ACTGGAACAA CACTCAACCC AATGAGCTG AATTAACGC GAATTTAAC TTTGTTTATT TTTCTAAATA CATTACAAT  
1821 ATTGGTTAAA CTTTTCGGG CATGAGACAA TAACCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
1891 TTAGGTGGCA CTTTTCGGG CATGAGACAA TAACCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
1961 TGATCCCGCT CCGTTCGCCC TATATCCCTT TTTTGGCGCA TTTTGCCTTC CACGAGTGGG TTACATCGAA CTGGATCTCA  
2031 TCAACATTTT CCGTTCGCCC TATATCCCTT TTTTGGCGCA TTTTGCCTTC CACGAGTGGG TTACATCGAA CTGGATCTCA  
2101 ACGCTGGTGA AAGTAAAGA GATCCTTGAG AGTTTTCGCC CCGAAGAAC CAGTTCGCTT TTTTGCCTTC CACGAGTGGG  
2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAAC CAGTTCGCTT TTTTGCCTTC CACGAGTGGG TTACATCGAA  
2241 GCTATGTGGC GCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT CCGGTATTAT  
2311 CAGAAATGACT TGGTTGAGTA CTCACCAAGT ACAGTGGTGA ACAGTGGTGA ACAGTGGTGA ACAGTGGTGA  
2381 TATGCAGTGC TGGCATAACC ACCGCTTTT CCATACCAAA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA  
2451 GAAGGAGCTA ACCGCTTTT CCATACCAAA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA  
2521 CTGAATGAAG CCATACCAAA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA  
2591 AACTATTAACT TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA  
2661 AGTTGCAGGA CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC  
2731 GAGCGTGGGT CTCGCGGTAT CATTCAGCA CATTCAGCA CATTCAGCA CATTCAGCA CATTCAGCA CATTCAGCA  
2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGTTCGCT GTAGTTCGCT GTAGTTCGCT GTAGTTCGCT  
2871 TAAGCATTTG TAACGTGTCAG ACCAAGTTTA CTCATATATA CTCATATATA CTCATATATA CTCATATATA CTCATATATA  
2941 TTTAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT TTCTTGAGAT TTCTTGAGAT TTCTTGAGAT  
3011 TCCACTGAGC GTACAGACCC GTACAGACCC GTACAGACCC GTACAGACCC GTACAGACCC GTACAGACCC GTACAGACCC  
3081 CTGCTGCTTG CAAACAAAAG AGGTAACCTGG CAAACAAAAG AGGTAACCTGG CAAACAAAAG AGGTAACCTGG CAAACAAAAG  
3151 CTTTTCCTGA AGGTAACCTGG CAAACAAAAG AGGTAACCTGG CAAACAAAAG AGGTAACCTGG CAAACAAAAG AGGTAACCTGG  
3221 TAGGCCACCA CTTCAAGAAC GTCTGAGCAC CGCTTACATA CCTCGCTCTG TCAAGACGAT TCAAGACGAT TCAAGACGAT  
3291 TGCTGCCAGT GGCGATAAGT GGCGATAAGT GGCGATAAGT GGCGATAAGT GGCGATAAGT GGCGATAAGT GGCGATAAGT  
3361 CGGTCGGGCT GAACGGGGG TCGTGCACCA CAGCCCAAGT CAGCCCAAGT CAGCCCAAGT CAGCCCAAGT CAGCCCAAGT  
3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CAGCTTCCCA GAGGTAACCG GAGGTAACCG GAGGTAACCG GAGGTAACCG  
3501 CGGCAGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGAGGTAACCG GGAGGTAACCG GGAGGTAACCG GGAGGTAACCG  
3571 GTCGGGTTT GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC GATGCTCGTC GATGCTCGTC GATGCTCGTC  
3641 AAAACGCCAG CAACGCGGCC TTTTACGGT TCCTGCGCTT TTGCTGGCCT TTGCTGGCCT TTGCTGGCCT TTGCTGGCCT  
3711 TGCGTTATCC CCTGATCTG TGGATAACCG TATTACCGCC TATTACCGCC TATTACCGCC TATTACCGCC TATTACCGCC  
3781 CGAACGACCG AGCGCAGCGA GTCACTGAGC GAGGAAGCGG AAGAGCGGCC AAGAGCGGCC AAGAGCGGCC AAGAGCGGCC  
3851 CCGCGCGTTG GCCGATTCTA TAATGCAGCT GGCACGACAG GTTTCCTCGAC GTTTCCTCGAC GTTTCCTCGAC GTTTCCTCGAC

FIG.\_39D

74 / 154

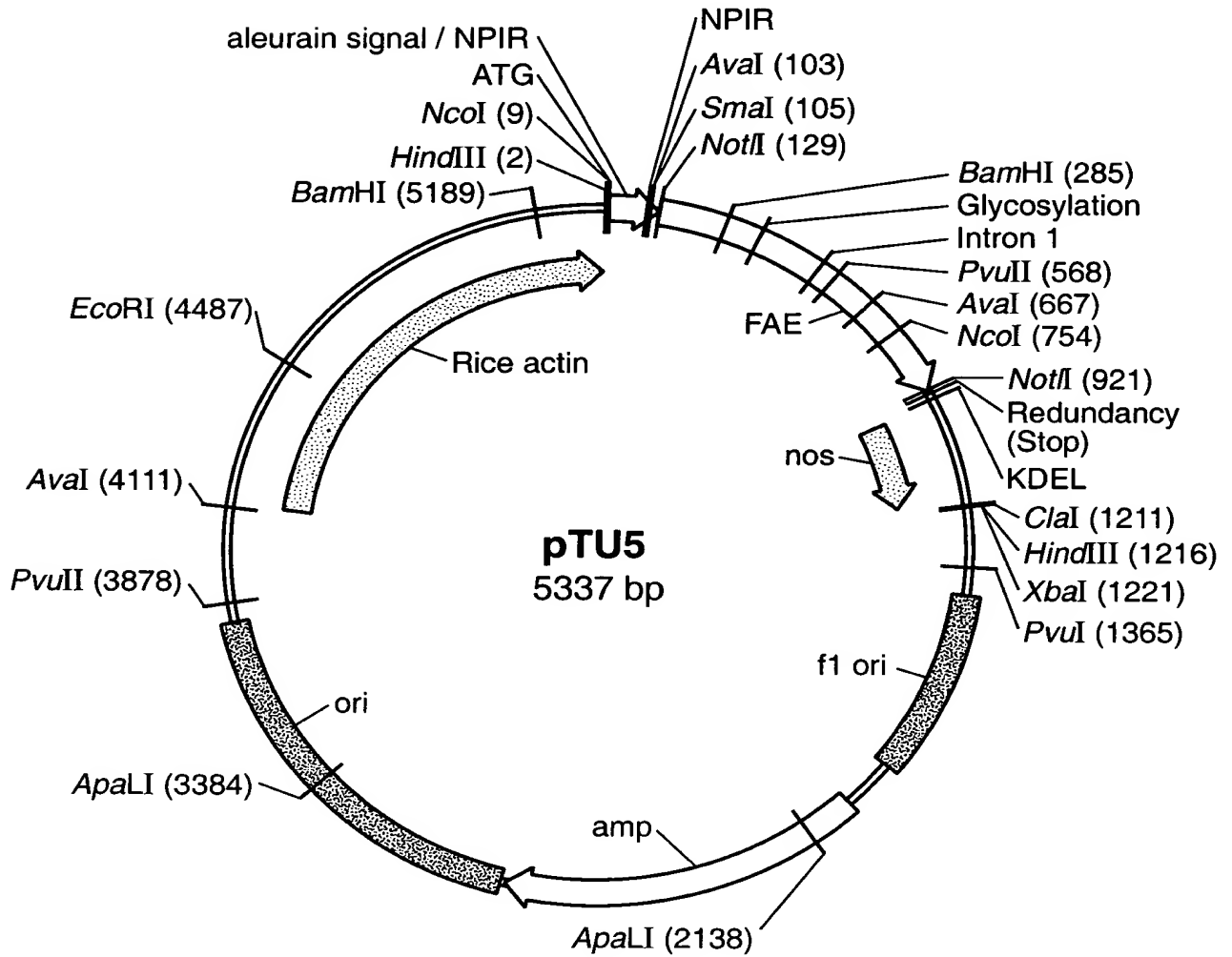
```

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC
                                     KpnI
                                     ~~~~~
4061 GCGCAATTAA CCTTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCCC CTCGAGGTCA TTCATATGCT
4131 TGAGAAGAGA GTCGGGATAG TCCAAAATAA AACAAAGTA AGATTACCTG GTCAAAAGTG AAAACATCAG
4201 TTAAGAGGTG GTATAAGTAA AATATCGGTA ATAAAGGTG GCCCAAAGTG AAATTACTC TTTTCTACTA
4271 TTATAAAAAT TGAGGATGTT TTGTCGGTAC TTTGATACGT CATTTTGTG TGAATTGGTT TTTAAGTTTA
4341 TTCGCGATTT GGAATGTCAT ATCTGTATTT GAGTCGGTTT TTAAGTTTCT TGTCTTGTG AATACAGAGG
4411 GATTGTGATA AGAAATATCT TTAATAAACCC CATATGCTAA TTTGACATAA TTTTGTGAGAA AAATATATAT
 EcorI
                                     ~~~~~
4481 TCAGGCGAAT TCCACAATGA ACAATAATAA GATTAAATA GCTTGCCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAA ATAAAGATA AACTTAGACT CAAAACATTT ACAAACAA AACCATAAAGT CCTAAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCCC AACCCACCCC AGTGCAGCCA
4691 ACTGGCAAAT AGTCTCCACC CCGGCACTA TCACCGTGAG TTGTCCGCAC CACCGCACGT CTCGCAGCCA
4761 AAAAAAAA AAGAAAGAAA AAAAAAGAAA AGAAACACAG CAGGTGGGTC CGGGTCGTGG GGGCCGGAAA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGGC CCGGCCCTCC CTCCGCTTCC AAAGAAACGC CCCCCATCGC
4901 CACTATATAC ATACCCCCC CTCTCCTCC ATCCCCC CCA ACCACCAACC ACCACCAACC CCACCTCCCTC
4971 CCCCCTCGCT GCCGACGAC GAGCTCCTCC CCCCCTCC CCGGCGCGCG TCCGCGCGCG CCGGTAACCA CCCCCCCCC
5041 CTCCTCTTTC TTCTCCGTT TTTTCTTTCG TCCTCGCTC GATCTTTGGC GATCTTTGGC CTTGGTAGTT TGGGTGGGCG
5111 AGAGCGGCTT CGTCGCCCAG ATCGGTGCGC GGGAGGGGCG GGATCTCGCG GCTGCGGTCT CCGGGCGGTGA
                                     BamHI
                                     ~~~~~
5181 GTCGGCCCCG ATCCTCGCGG GGAATGGGCG TCCTCGGATGT AGATCTTCTT TCTTCTTCT TTTTGTGGTA
5251 GAAATTGAAT CCTCAGCAT TGTTTCATCGG TAGTTTTTCT TTTTCATGATT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTTGTAGC

```

FIG.\_39E

75 / 154



**FIG.\_40A**

76 / 154

```

HindIII NcoI
~~~~~
1  AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT
   TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGGCACGA

51  GGCCACGGCC GCCGTCGCCG TCGCCTCCTC CTCCTCCTTC GCCGACTCCA
   CCGGTGCCGG CGGCAGCGGC AGCGGAGGAG GAGGAGGAAG CGGCTGAGGT

      SmaI
      ~~~~~
 AvaI
      ~~~~~
                                NotI
                                ~~~~~
101 ACCCGGGCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC GCAGGGCATC
 TGGGCCCGGC CGGGCAGTGG CTGGCGCGCC GCGGAGGTG CGTCCCGTAG

151 TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC
 AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTCC

201 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA
 ACGGATGCGG CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCTCTCT

 BamHI
                                ~~~~~
251  AAATTTACAA TTCTCAAAC TACATTAACG GATGGATCCT CCGCGACGAC
     TTTAAATGTT AAGAGTTTGA CTGTAATTGC CTACCTAGGA GGCCTGCTG

301  AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCAC TGGTA GTGATACGAA
     TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT

351  TCTACAAC TC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC
     AGATGTTGAG CTATGATTGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG

401  AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC
     TTACGTTGCC AACACTTCAT GTGCCACCTA TAATATAACC TACCCAGAGG

451  GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA GCCAGTATCC
     CAGGTCTTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT CGGTCATAGG

501  GGA CTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG
     CCTGATGCGC GACTGGCACT GGCCGGTGMG GGAGCCGCGG AGGGACCGCC

                                PvuII
                                ~~~~~
551 CACTCACTGC CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC
 GTGAGTGACG GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GGC GGACATG

601 ACCTTCGGCG AACCGCGCAG CGGCAATCAG GCCTTCGCGT CGTACATGAA
 TGG AAGCCGC TTGGCGCGTC GCCGTTAGTC CGGAAGCGCA GCATGTACTT

 AvaI
                                ~~~~~
651  CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA
     GCTACGGAAG GTTCGGAGCT CGGGTCTATG CTGCGTCATA AAGGCCAGT

```

**FIG.\_40B**



77 / 154

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC  
GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG

NcoI

~~~~~

751 GCCCATGGCG GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCCAGAA
CGGGTACCGC CACATCTCAT GACCTCGCAA CTAGGAATGT CGCGGGTCTT

801 CACATTTGTC TGCAGTGGGG ATGAAGTGCA GTGCTGTGAG GCCCAGGGCG
GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC CGGGTCCCGC

851 GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC
CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAAACCCTA CTGCTCGCCG

NotI

~~~~~

901 GCATGCACCT GGCCGGTCCG GGCCGCGGAA CCACTGAAGG ATGAGCTGTA  
CGTACGTGGA CCGGCCAGCG CCGGCGCCTT GGTGACTTCC TACTCGACAT

951 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT  
TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA

1001 CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT  
GGACAACGGC CAGAACGCTA CTAATAGTAT ATTAAAGACA ACTTAATGCA

1051 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG  
ATTCGTACAT TATTAATTGT ACATTACGTA CTGCAATAAA TACTCTACCC

1101 TTTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC  
AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG

1151 AAAATATAGC GCGCAAACCTA GGATAAATTA TCGCGCGCGG TGTATCTAT  
TTTTATATCG CGCGTTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA

XbaI

~~~~~

ClaI HindIII

~~~~~

1201 GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC  
CAATGATCTA GCTATTCGAA GATCTCGCCG GCCACCTCGA GGTTAAGCGG

1251 CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT TTACAACGTC  
GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG

1301 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT  
CACTGACCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA

PvuII

~~~~~

1351 CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
GGGGGAAAGC GGTCGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG

1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG
AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCCTGCGC GGGACATCGC

FIG. 40C

78 / 154

| | | | | | |
|----------------|--------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| 1451 | GCGCATTAAG
CGCGTAATTC | CGCGGCGGGT
GCGCCGCCCA | GTGGTGGTTA
CACCACCAAT | CGCGCAGCGT
GCGCGTCGCA | GACCGCTACA
CTGGCGATGT |
| 1501 | CTTGCCAGCG
GAACGGTCGC | CCCTAGCGCC
GGGATCGCGG | CGCTCCTTTC
GCGAGGAAAG | GCTTTCTTCC
CGAAAGAAGG | CTTCCTTTCT
GAAGGAAAGA |
| 1551 | CGCCACGTTT
GCGGTGCAAG | GCCGGCTTTC
CGGCCGAAAG | CCCGTCAAGC
GGGCAGTTTC | TCTAAATCGG
AGATTTAGCC | GGGCTCCCTT
CCCGAGGGAA |
| 1601 | TAGGGTTCCG
ATCCCAAGGC | ATTTAGTGCT
TAAATCACGA | TTACGGCACC
AATGCCGTGG | TCGACCCCAA
AGCTGGGGTT | AAAAC TTGAT
TTTTGAACTA |
| 1651 | TAGGGTGATG
ATCCCACTAC | G TTCACGTAG
CAAGTGCATC | TGGGCCATCG
ACCCGGTAGC | CCCTGATAGA
GGGACTATCT | CGGTTTTTTCG
GCCAAAAAGC |
| 1701 | CCCTTTGACG
GGGAAACTGC | TTGGAGTCCA
AACCTCAGGT | CGTTCTTTAA
GCAAGAAATT | TAGTGGACTC
ATCACCTGAG | TTGTTCCAAA
AACAAGGTTT |
| 1751 | CTGGAACAAC
GACCTTGTTG | ACTCAACCCT
TGAGTTGGGA | ATCTCGGTCT
TAGAGCCAGA | ATTCTTTTGA
TAAGAAAAC T | TTTATAAGGG
AAATATTCCC |
| 1801 | ATTTTGCCGA
TAAAACGGCT | TTTCGGCCTA
AAAGCCGGAT | TTGGTTAAAA
AACCAATTTT | AATGAGCTGA
T TACTCGACT | TTTAACAAAA
AAATTGTTTT |
| 1851 | ATTTAACGCG
TAAATTGCGC | AATTTTAAAC
TTAAAATTGT | AAATATTAAC
TTTATAATTG | GCTTACAATT
CGAATGT TAA | TAGGTGGCAC
ATCCACCGTG |
| 1901 | TTTTCGGGGA
AAAAGCCCCT | AATGTGCGCG
TTACACGCGC | GAACCCCTAT
CTTGGGGATA | TTGTTTATTT
AACAAATAAA | TTCTAAATAC
AAGATTTATG |
| 1951 | ATTCAAATAT
TAAGTTTATA | GTATCCGCTC
CATAGGCGAG | ATGAGACAAT
TACTCTGTTA | AACCCTGATA
TTGGGACTAT | AATGCTTCAA
TTACGAAGTT |
| 2001 | TAATATTGAA
ATTATAACTT | AAAGGAAGAG
TTTCCTTCTC | TATGAGTATT
ATACTCATAA | CAACATTTCC
GTTGTAAAGG | GTGTCGCCCT
CACAGCGGGA |
| 2051 | TATTCCTTTT
ATAAGGGAAA | TTTGCGGCAT
AAACGCCGTA | TTTGCCTTCC
AAACGGAAGG | TGTTTTTGCT
ACAAAAACGA | CACCCAGAAA
GTGGGTCTTT |
| ApaLI
~~~~~ | | | | | |
| 2101 | CGCTGGTGAA
GCGACCACTT | AGTAAAAGAT
TCATTTTCTA | GCTGAAGATC
CGACTTCTAG | AGTTGGGTGC
TCAACCCACG | ACGAGTGGGT
TGCTCACCCA |
| 2151 | TACATCGAAC
ATGTAGCTTG | TGGATCTCAA
ACCTAGAGTT | CAGCGGTAAG
GTCGCCATTC | ATCCTTGAGA
TAGGAACTCT | GTTTTCGCCC
CAAAAGCGGG |
| 2201 | CGAAGAACGT
GCTTCTTGCA | TTTCCAATGA
AAAGGTTACT | TGAGCACTTT
ACTCGTGAAA | TAAAGTTCTG
ATTTCAAGAC | CTATGTGGCG
GATACACCGC |
| 2251 | CGGTATTATC
GCCATAATAG | CCGTATTGAC
GGCATAACTG | GCCGGGCAAG
CGGCCCCGTT | AGCAACTCGG
TCGTTGAGCC | TCGCCGCATA
AGCGGCGTAT |
| 2301 | CACTATTCTC
GTGATAAGAG | AGAATGACTT
TCTTACTGAA | GGTTGAGTAC
CCAAC TCATG | TCACCAGTCA
AGTGGTCAGT | CAGAAAAGCA
GTCTTTTCGT |

FIG._40D

79 / 154

| | | | | | |
|------|---------------------------|--------------------------|--------------------------|---------------------------|--------------------------|
| 2351 | TCTTACGGAT
AGAATGCCTA | GGCATGACAG
CCGTACTGTC | TAAGAGAATT
ATTCTCTTAA | ATGCAGTGCT
TACGTCACGA | GCCATAACCA
CGGTATTGGT |
| 2401 | TGAGTGATAA
ACTCACTATT | CACTGCGGCC
GTGACGCCGG | AACTTACTTC
TTGAATGAAG | TGACAACGAT
ACTGTTGCTA | CGGAGGACCG
GCCTCCTGGC |
| 2451 | AAGGAGCTAA
TTCCTCGATT | CCGCTTTTTT
GGCGAAAAAA | GCACAACATG
CGTGTGTGAC | GGGGATCATG
CCCCTAGTAC | TAACTCGCCT
ATTGAGCGGA |
| 2501 | TGATCGTTGG
ACTAGCAACC | GAACCGGAGC
CTTGGCCTCG | TGAATGAAGC
ACTTACTTCG | CATACCAAAC
GTATGGTTTG | GACGAGCGTG
CTGCTCGCAC |
| 2551 | ACACCACGAT
TGTGGTGCTA | GCCTGTAGCA
CGGACATCGT | ATGGCAACAA
TACCGTTGTT | CGTTGCGCAA
GCAACGCGTT | ACTATTAAC
TGATAATTGA |
| 2601 | GGCGAACTAC
CCGCTTGATG | TTACTCTAGC
AATGAGATCG | TTCCCGGCAA
AAGGGCCGTT | CAATTAATAG
GTTAATTATC | ACTGGATGGA
TGACCTACCT |
| 2651 | GGCGGATAAA
CCGCCTATTT | GTTGCAGGAC
CAACGTCTTG | CACTTCTGCG
GTGAAGACGC | CTCGGCCCTT
GAGCCGGGAA | CCGGCTGGCT
GGCCGACCGA |
| 2701 | GGTTTATTGC
CCAAATAACG | TGATAAATCT
ACTATTTAGA | GGAGCCGGTG
CCTCGGCCAC | AGCGTGGGTC
TCGCACCCAG | TCGCGGTATC
AGCGCCATAG |
| 2751 | ATTGCAGCAC
TAACGTCGTG | TGGGGCCAGA
ACCCCGGTCT | TGGTAAGCCC
ACCATTGCGG | TCCCGTATCG
AGGGCATAGC | TAGTTATCTA
ATCAATAGAT |
| 2801 | CACGACGGGG
GTGCTGCCCC | AGTCAGGCAA
TCAGTCCGTT | CTATGGATGA
GATACCTACT | ACGAAATAGA
TGCTTTTATCT | CAGATCGCTG
GTCTAGCGAC |
| 2851 | AGATAGGTGC
TCTATCCACG | CTCACTGATT
GAGTGACTAA | AAGCATTGGT
TTCGTAACCA | AACTGTCAGA
TTGACAGTCT | CCAAGTTTAC
GGTTCAAATG |
| 2901 | TCATATATAC
AGTATATATG | TTTAGATTGA
AAATCTAACT | TTTAAACTTT
AAATTTTGAA | CATTTTTTAAT
GTAAAAATTA | TTAAAAGGAT
AATTTTCCTA |
| 2951 | CTAGGTGAAG
GATCCACTTC | ATCCTTTTTG
TAGGAAAAAC | ATAATCTCAT
TATTAGAGTA | GACCAAATC
CTGGTTTTAG | CCTTAACGTG
GGAATTGCAC |
| 3001 | AGTTTTTCGTT
TCAAAAGCAA | CCACTGAGCG
GGTGACTCGC | TCAGACCCCG
AGTCTGGGGC | TAGAAAAGAT
ATCTTTTCTA | CAAAGGATCT
GTTTCCTAGA |
| 3051 | TCTTGAGATC
AGAACTCTAG | CTTTTTTTCT
GAAAAAAGA | GCGCGTAATC
CGCGCATTAG | TGCTGCTTGC
ACGACGAACG | AAACAAAAAA
TTTGTTTTTT |
| 3101 | ACCACCGCTA
TGGTGGCGAT | CCAGCGGTGG
GGTCGCCACC | TTTGTGTGCC
AAACAAACGG | GGATCAAGAG
CCTAGTTCTC | CTACCAACTC
GATGGTTGAG |
| 3151 | TTTTTCCGAA
AAAAAGGCTT | GGTAACTGGC
CCATTGACCG | TTCAGCAGAG
AAGTCGTCTC | CGCAGATACC
GCGTCTATGG | AAATACTGTC
TTTATGACAG |
| 3201 | CTTCTAGTGT
GAAGATCACA | AGCCGTAGTT
TCGGCATCAA | AGGCCACCAC
TCCGGTGGTG | TTCAAGAACT
AAGTTCTTGA | CTGTAGCACC
GACATCGTGG |
| 3251 | GCCTACATAC
CGGATGTATG | CTCGCTCTGC
GAGCGAGACG | TAATCCTGTT
ATTAGGACAA | ACCAGTGGCT
TGGTCACCGA | GCTGCCAGTG
CGACGGTCAC |

FIG. 40E

80 / 154

3301 GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT
CGCTATTTCAG CACAGAATGG CCCAACCTGA GTTCTGCTAT CAATGGCCTA

ApaLI

~~~~~

3351 AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGACAC AGCCCAGCTT  
TTCCGCGTCG CCAGCCCGAC TTGCCCCCA AGCACGTGTG TCGGGTCGAA

3401 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG  
CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTGCA CTCGATACTC

3451 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC  
TTTCGCGGTG CGAAGGGCTT CCCTCTTTCC GCCTGTCCAT AGGCCATTGC

3501 GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC  
CCGTCCAGC CTTGTCCTCT CGCGTGCTCC CTCGAAGGTC CCCCTTTGCG

3551 CTGGTATCTT TATAGTCTTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC  
GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACTCGCAG

3601 GATTTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC  
CTAAAACAC TACGAGCAGT CCCCCGCCT CGGATACCTT TTTGCGGTGC

3651 AACGCGGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT  
TTGCGCCGGA AAAATGCCAA GGACCGGAAA ACGACCGGAA AACGAGTGTA

3701 GTTCTTTCCT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT  
CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA

3751 TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG  
AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTGCTGGCT CGCGTCGCTC

3801 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC CGCCTCTCCC  
AGTCACTCGC TCCTTCGCCT TCTCGCGGGT TATGCGTTTG GCGGAGAGGG

PvuII

~~~~~

3851 CGCGCGTTGG CCGATTCATT AATGCAGCTG GCACGACAGG TTTCCCGACT
GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGCTGTCC AAAGGGCTGA

3901 GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT
CCTTTCGCCC GTCACCTCGC TTGCGTTAAT TACACTCAAT CGAGTGAGTA

3951 TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGG
ATCCGTGGGG TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC

4001 AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT
TTAACACTCG CCTATTGTGA AAGTGTGTCC TTTGTCGATA CTGGTACTAA

4051 ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGGAACAAAA GCTGGGTACC
TGCGGTTCGC GCGTTAATTG GGAGTGATTT CCCTTGTTTT CGACCCATGG

AvaI

~~~~~

4101 GGGCCCCCCC TCGAGGTCAT TCATATGCTT GAGAAGAGAG TCGGGATAGT  
CCCGGGGGGG AGCTCCAGTA AGTATACGAA CTCTTCTCTC AGCCCTATCA

**FIG. 40F**

81 / 154

|                |                          |                          |                          |                          |                          |
|----------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 4151           | CCAAAATAAA<br>GGTTTTATTT | ACAAAGGTAA<br>TGTTTCCATT | GATTACCTGG<br>CTAATGGACC | TCAAAAGTGA<br>AGTTTTCACT | AAACATCAGT<br>TTTGTAGTCA |
| 4201           | TAAAAGGTGG<br>ATTTTCCACC | TATAAGTAAA<br>ATATTCATTT | ATATCGGTAA<br>TATAGCCATT | TAAAAGGTGG<br>ATTTTCCACC | CCCAAAGTGA<br>GGGTTTCACT |
| 4251           | AATTTACTCT<br>TTAAATGAGA | TTTCTACTAT<br>AAAGATGATA | TATAAAAATT<br>ATATTTTTAA | GAGGATGTTT<br>CTCCTACAAA | TGTCGGTACT<br>ACAGCCATGA |
| 4301           | TTGATACGTC<br>AACTATGCAG | ATTTTTGTAT<br>TAAAAACATA | GAATTGGTTT<br>CTTAACCAAA | TTAAGTTTAT<br>AATTCAAATA | TCGCGATTG<br>AGCGCTAAAC  |
| 4351           | GAAATGCATA<br>CTTTACGTAT | TCTGTATTTG<br>AGACATAAAC | AGTCGGTTTT<br>TCAGCCAAAA | TAAGTTTCGT<br>ATTCAAGCAA | GCTTTTGTAA<br>CGAAAACATT |
| 4401           | ATACAGAGGG<br>TATGTCTCCC | ATTTGTATAA<br>TAAACATATT | GAAATATCTT<br>CTTTATAGAA | TAAAAAACCC<br>ATTTTTTGGG | ATATGCTAAT<br>TATACGATTA |
| EcoRI<br>~~~~~ |                          |                          |                          |                          |                          |
| 4451           | TTGACATAAT<br>AACTGTATTA | TTTTGAGAAA<br>AAAACCTCTT | AATATATATT<br>TTATATATAA | CAGGCGAATT<br>GTCCGCTTAA | CCACAATGAA<br>GGTGTTACTT |
| 4501           | CAATAATAAG<br>GTTATTATTC | ATTAAAATAG<br>TAATTTTATC | CTTGCCCCCG<br>GAACGGGGGC | TTGCAGCGAT<br>AACGTCGCTA | GGGTATTTTT<br>CCCATAAAAA |
| 4551           | TCTAGTAAAA<br>AGATCATTTT | TAAAAGATAA<br>ATTTTCTATT | ACTTAGACTC<br>TGAATCTGAG | AAAACATTTA<br>TTTTGTAAAT | CAAAAACAAC<br>GTTTTTGTG  |
| 4601           | CCCTAAAGTC<br>GGGATTTTCA | CTAAAGCCCA<br>GATTTTCGGT | AAGTGCTATG<br>TTCACGATAC | CACGATCCAT<br>GTGCTAGGTA | AGCAAGCCCA<br>TCGTTCGGGT |
| 4651           | GCCCAACCCA<br>CGGGTTGGGT | ACCCAACCCA<br>TGGGTGTTGG | ACCCACCCCA<br>TGGGTGTTGG | GTGCAGCCAA<br>CACGTCGGTT | CTGGCAAATA<br>GACCGTTTAT |
| 4701           | GTCTCCACCC<br>CAGAGGTGGG | CCGGCACTAT<br>GGCCGTGATA | CACCGTGAGT<br>GTGGCACTCA | TGTCCGCACC<br>ACAGGCGTGG | ACCGCACGTC<br>TGGCGTGCAG |
| 4751           | TCGCAGCCAA<br>AGCGTCGGTT | AAAAAAAAAA<br>TTTTTTTTTT | AGAAAGAAAA<br>TCTTCTTTTT | AAAAGAAAAA<br>TTTTCTTTTT | GAAAAACAGC<br>CTTTTTGTGC |
| 4801           | AGGTGGGTCC<br>TCCACCCAGG | GGGTCGTGGG<br>CCCAGCACCC | GGCCGGAAAA<br>CCGGCCTTTT | GCGAGGAGGA<br>CGCTCCTCCT | TCGCGAGCAG<br>AGCGCTCGTC |
| 4851           | CGACGAGGCC<br>GCTGCTCCGG | CGGCCCTCCC<br>GCCGGGAGGG | TCCGCTTCCA<br>AGGCGAAGGT | AAGAAACGCC<br>TTCCTTTGCG | CCCCATCGCC<br>GGGGTAGCGG |
| 4901           | ACTATATACA<br>TGATATATGT | TACCCCCCCC<br>ATGGGGGGGG | TCTCCTCCCA<br>AGAGGAGGGT | TCCCCCAAC<br>AGGGGGGTTG  | CCTACCACCA<br>GGATGGTGGT |
| 4951           | CCACCACCAC<br>GGTGGTGGTG | CACCTCCTCC<br>GTGGAGGAGG | CCCCTCGCTG<br>GGGGAGCGAC | CCGGACGACG<br>GGCCTGCTGC | AGCTCCTCCC<br>TCGAGGAGGG |
| 5001           | CCCTCCCCCT<br>GGGAGGGGGA | CCGCCGCCGC<br>GGCGGCGGCG | CGGTAACCAC<br>GCCATTGGTG | CCCGCCCCCT<br>GGGCGGGGAG | TCCTCTTTCT<br>AGGAGAAAGA |

**FIG.\_40G**

82 / 154

5051 TTCTCCGTTT TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT  
AAGAGGCAAA AAAAAAAGCA GAGCCAGAGC TAGAAACCGG AACCATCAAA

5101 GGGTGGGCGA GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG  
CCCACCCGCT CTCGCCGAAG CAGCGGGTCT AGCCACGCGC CCTCCCCGCC

BamHI

~~~~~

5151 GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG
CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCCT AGGAGCGCCC

5201 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG
CTTACCCCGA GAGCCTACAT CTAGAAGAAA GAAAGAAGAA AACACCATC

5251 AATTTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT
TTAAACTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAGAA AAGTACTAAA

5301 GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC
CACTGTTTAC GTCGGAGCAC GCCTCGAAAA AACATCG

FIG._40H

83 / 154

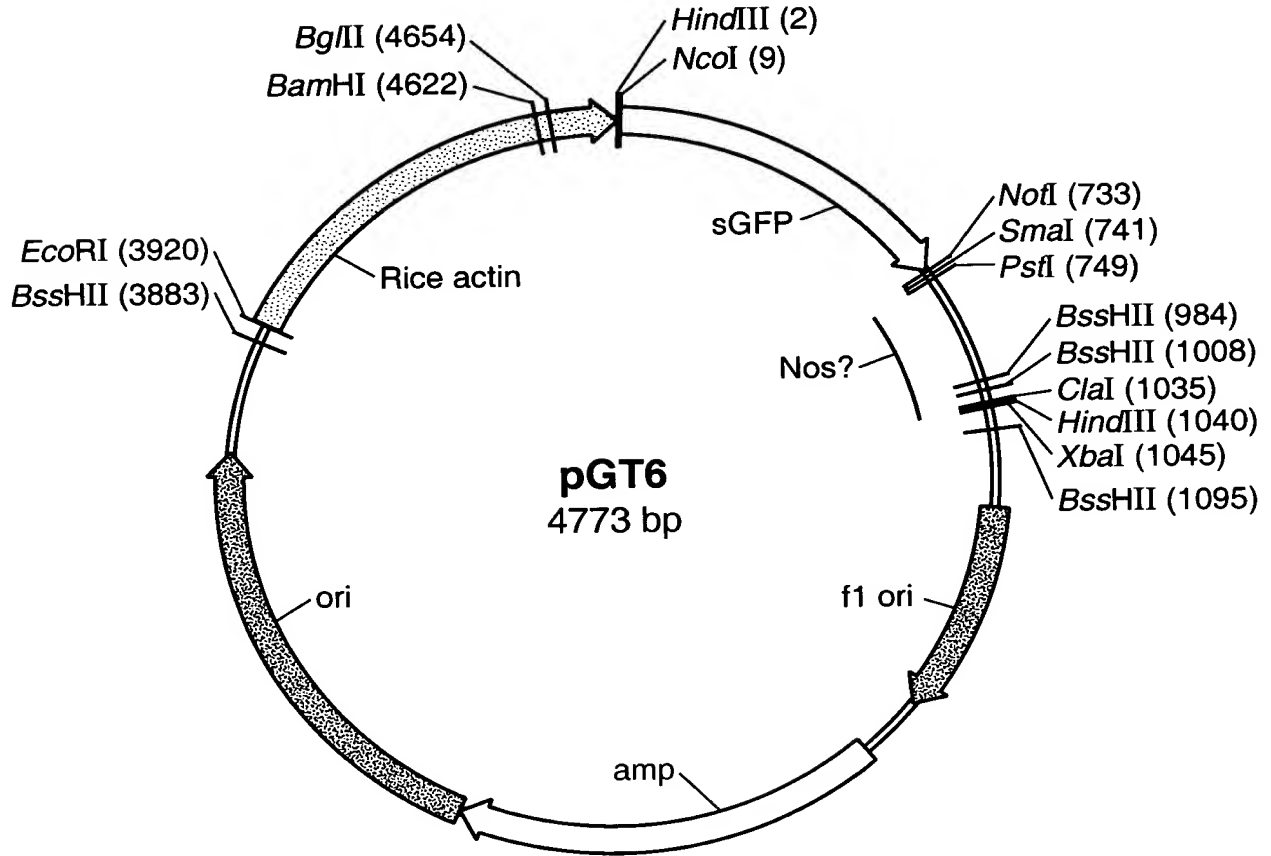


FIG._41A

+

HindIII NcoI
~~~~~  
1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCG GGGTGGTGCC CATCCTGGTC GAGCTGGACG  
TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAAGTGGC CCCACCACGG GTAGGACCAG CTCGACCTGC  
71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT GCCACCTACG GCAAGCTGAC  
CGCTGCACTT GCCGGTGTTC AAGTCGCACA GCGCGCTCCC GCTCCCGCTA CCGTGGATGC CGTTCGACTG  
141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCCAACC TCGTGACCAC CTTACACCTAC  
GGACTTCAAG TAGACGTGGT GGCCGTTCGA CCGGCACGGG ACCGGGTGGG AGCACTGGTG GAAGTGGATG  
211 GCGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT CTTCAAAGTCC GCCATGCCCG  
CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGTGAA GAAGTTCAGG CCGTACGGGC  
281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCCGC CCGAGGTGAA  
TTCCGATGCA GGTCCCTCGC TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT  
351 GTTCGAGGGC GACACCCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC  
CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTTGTAG  
421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAAG CAGAAGAACG  
GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTT GTCTTCTTGC  
491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG CAGCTCGCCG ACCACTACCA  
CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGATGGT  
561 GCAGAACACC CCCATCGGGC ACGGCCCCGT GCTGCTGCCC GACAACCACT ACCTGAGCAC CCAGTCCGCC  
CGTCTTGTGG GGGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA TGGACTCGTG GGTACGGCGG  
631 CTGAGCAAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA  
GACTCGTTTC TGGGGTTGCT CTTCCGCGCTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

FIG..41B

+



85 / 154

```

SmaI
~~~~~
 NotI PstI
~~~~~
701 CTCACGGCAT GGACGAGCTG TACAAGTAA GCGCCGCC CCGCTGCAGG GAAACCACTG AAGGATGAGC
    GAGTGCCGTA CCGTCTCGAC ATGTTCAATT CGCCGGCGGG CCGACGTC CTTTGGTGAC TTCTTACTCG

771 TGTAAGAAG CAGATCGTTC AAACATTTGG CAATAAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG
    ACATTTCTTC GTCTAGCAAG TTTGTAAACC GTTATTTCAG AGAATTCTAA CTTAGGACAA CGGCCAGAAC

841 CGATGATTAT CATATAATT CTGTTGAATT ACGTTAAGCA TGTAATAAAT AACATGTAAT GCATGACGTT
    GCTACTAATA GTATATTAA GACAACTTAA TGCAATTCGT ACATTATTAA TTGTACATTA CGTACTGCAA

911 ATTTATGAGA TGGGTTTTTA TGATTAGAGT CCGGCAATTA TACATTTAAT ACGCGATAGA AAACAAATA
    TAAATACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAT ATGTAAATTA TGCCTATCT TTTGTTTTAT

      XbaI
      ~~~~~

 ClaI HindIII
      ~~~~~
981 TAGCGCGCAA ACTAGGATAA ATTATCGCGC GCGGTGTCAT CTATGTTACT AGATCGATAA GTTCTAGAG
    ATCGCGCGTT TGATCCTATT TAATAGCGCG CGCCACAGTA GATACAATGA TCTAGCTATT CGAAGATCTC

      BssHII
      ~~~~~
1051 CGGCCGGTGG AGCTCCAATT CGCCCTATAG TGAGTCGTAT TACGCGCGCT CACTGGCCGT CGTTTACAA
 GCCGGCCACC TCGAGGTTAA GCGGGATATC ACTCAGCATA ATGCGCGCGA GTGACCGGCA GCAAAATGTT

1121 CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTTAATC GCCTTGAGC ACATCCCCCT TTCGCCAGCT
 GCAGCACTGA CCTTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG TGTAGGGGGA AAGCGGTCGA

1191 GCGGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGGA
 CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAGGGT TGTCAACGCG TCGGACTTAC CGCTTACCTT

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGCGG GGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC
 GCGCGGACA TCGCCGCGTA ATTGCGCGG CCCACACCAC CAATGCGCGT CGCACTGGCG ATGTGAACGG

```

FIG.\_41C

```

1331 AGCGCCCTAG CGCCCGCTCC TTTTCGCTTTC TTCCCTTCTT TCTCGCCAC GTTCGCCGC TTTCCCCGTC
TCGCGGGATC GCGGGCGAGG AAGCGAAAG AAGGAAGGA AAGAGCGGTG CAAGCGGCCG AAAGGGGCAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGT TCCGATTAG TGCTTTACGG CACCTCGACC CCAAAAAACT
TTGAGATT TT AGCCCCCGAG GGAATCCCA AGGCTAAATC ACGAAATGCC GTGGAGCTGG GGTTTTGTGA

1471 TGATTAGGT GATGGTTTAC GTAGTGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGACT ATCTGCCAA AAGCGGGA A CTGCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA CCCATATCTCG GTCTATTTCT
AGGTGCAAGA AATTATCACC TGAGAACAA GTTTGACCTT GTTGAGGTT GGGATAGAGC CAGATAAGAA

1611 TTGATTTATA AGGGATTTTG CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAAATTAA
AACTAAATAT TCCCTAAAC GGCTAAAGCC GGATAACCAA TTTTTTTACTC GACTAAATG TTTTAAAT

1681 CGCGAATTTT AACAAAAATAT TAACGCTTAC AATTAGGTG GCACTTTTCG GGGAAATGTG CGCGGAACCC
CGCTTTAAA TTGTTTTTATA ATTGCGAATG TTAATCCAC CGTGAAAAAGC CCCTTTTACAC CGCCTTGGG

1751 CTATTTGTTT ATTTTCTTAA ATACATTCAA ATATGTATCC GTCATGAGA CAATAACCCCT GATAAATGCT
GATAAACAAA TAAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTACGA

1821 TCAATAATAT TGA AAAAAGGA AGAGTATGAG TATTCACAT TCCCGTGTG CCCTTATTCC CTTTTTTCG
AGTTATTATA ACTTTTTCCT TCTCATACTC ATAAGTTGTA AAGGCACAGC GGGAAATAAGG GAAAAACGC

1891 GCATTTTGCC TTCTGTGTTT TGCTCACCCA GAAACGCTGG TGAAGTAAA AGATGCTGAA GATCAGTTGG
CGTAAACGG AAGGACAAA ACGAGTGGGT CTTTGCAGC ACTTTCATTT TCTACGACTT CTAGTCAACC

1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAAGATCCTT GAGAGTTTTC GCCCGAAGA
CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTGCGC ATTCTAGGAA CTCTCAAAAG CGGGGCTTCT

2031 ACGTTTCCA ATGATGAGCA CTTTAAAGT TCTGCTATGT GCGCGGGTAT TATCCCGTAT TGACGCCGGG
TGCAAAAAGGT TACTACTCGT GAAAATTTC AAGACGATACA CCGCGCCATA ATAGGCATA ACTGCGGCC

2101 CAAGAGCAAC TCGGTCGCC CATACACTAT TCTCAGAATG ACTTGGTTGA GTACTCACCA GTCACAGAAA
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

```

FIG.-41D

87 / 154

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC  
TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG

2241 GGCCAACTTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGGAT  
CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGCGAA AAAACGTGTT GTACCCCTTA

2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA  
GTACATTGAG CGGAAC TAGC AACCCTTGGC CTCGACTTAC TTCGGTATGG TTTGCTGCTC GCACTGTGGT

2381 CGATGCCTGT AGCAATGGCA ACAACGTTGC GCAAACCTATT AACTGGCGAA CTACTTACTC TAGCTTCCCG  
GCTACGGACA TCGTTACCGT TGTGCAACG CGTTTGATAA TTGACCGCTT GATGAATGAG ATCGAAGGCG

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGCGCTCGGC CCTTCCGGCT  
CGTTGTTAAT TATCTGACCT TGTGCAACG ACCTCCGCTT ATTTCAACGT CCTGGTGAAG ACGGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG TATCATTTGCA GCACTGGGGC  
CCGACCAAAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA  
GTCTACCATT CGGGAGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCCCTCACT GATTAAGCAT TGGTAACTGT CAGACCAAGT TTACTCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTGCTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTTTAA ACTTCATTTT TAATTTTAAA GGATCTAGGT GAAGATCCCT TTTGATAAATC  
TATGAAATCT AACTAAATTT TGAAGTAAAA ATTAATTTT CCTAGATCCA CTTCTAGGAA AAACATATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG  
AGTACTGGTT TTAGGGAATT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAAACAA AAAAACCCACC GCTACCAGCG  
TAGAAGAACT CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGTG TTTTGTGGTG CGATGGTCCG

2941 GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA  
CACCAACAA ACGGCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAGTCG TCTCGCGTCT

FIG. 41E

88 / 154

```

3011 TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG CACCGCCTAC
 ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG
 TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCAAC

3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGGTCCG GCTGAACGGG GGGTTCGTGC ACACAGCCCA
 CTGAGTTCTG CTATCAATGG CCTATTCCGC GTGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTCGGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGGCTTC
 CGAACCTCGC TTGCTGGATG TGGCTTGAAT TATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT
 GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTG CTCTCGCGTG CTCCCTCGAA

3361 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT
 GGTCCCCCTT TGCGGACCAT AGAATATCA GGACAGCCCA AAGCGGTGA GACTGAAC TC GACGCTAAA

3431 TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTAC GGTTCCTGGC
 ACACACTAGG CAGTCCCCC GCCTCGGATA CCTTTTGGC GTCGTGCGC CGAAAAATG CCAAGGACCG

3501 CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCTGATT TCGTGGATA CCGTATTACC
 GAAAAACGACC GGAAAAACGAG TGTACAAGAA AGGACGCAAT AGGGACTAA GACACCTATT GGCAATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG
 CGGAAACTCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGCGTC GCTCAGTCAC TCGCTCCTTC

3641 CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCGCGCG TTGGCCGATT CATTAATGCA GCTGGCACGA
 GCCTTCTCGC GGGTTATGCG TTTGGCGGAG AGGGCGCGC AACCGGCTAA GTAATTACGT CGACCGTGCT

3711 CAGGTTTCCC GACTGGAAAG CGGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTAGGCA
 GTCCAAAGGG CTGACCTTTC GCCCGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT

3781 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATA CAATTTCACA
 GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCCCTATT GTTAAAGTGT

```

FIG.\_41F

**FIG. 41G**

```

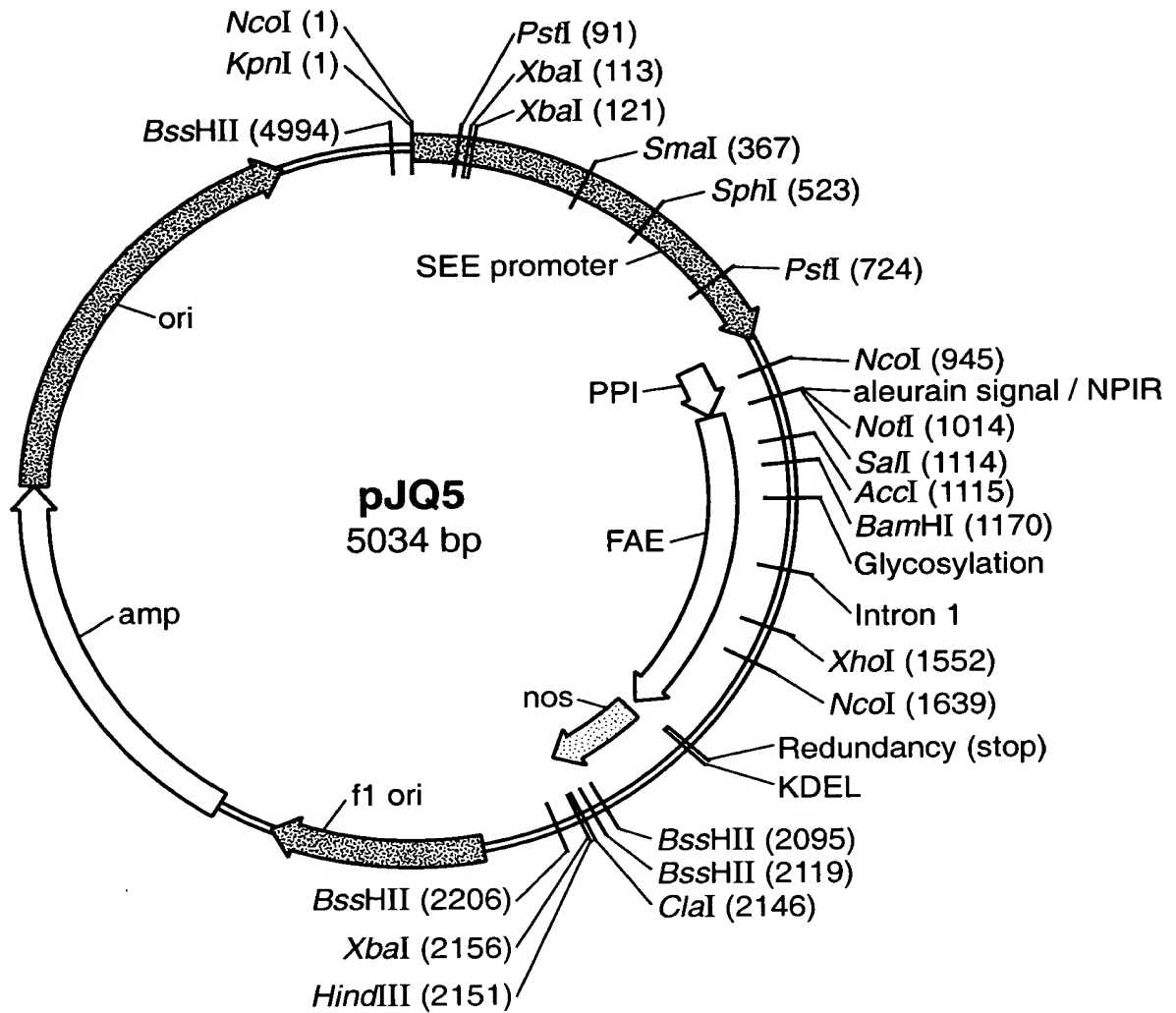
 BamHI BglII
      ~~~~~                               ~~~~~
4621  GGATCCTCGC  GGGGAATGGG  GCTCTCGGAT  GTAGATCTTC  TTTCTTTCTT  CTTTTTGTGG  TAGAATTGA
      CCTAGGAGCG  CCCCTTACCC  CGAGAGCCCTA  CATCTAGAAG  AAAGAAAGAA  GAAAAACACC  ATCTTAACT

4691  ATCCCCTCAGC  ATTGTTTCATC  GGTAGTTTTT  CTTTTCATGA  TTTGTGACAA  ATGCAGCCTC  GTGCCGGAGCT
      TAGGGAGTCG  TAACAAGTAG  CCATCAAAA  GAAAAGTACT  AAACACTGTT  TACGTCGGAG  CACGCCCTCGA

4761  TTTTGTAGG  TAG
      AAAAACATCC  ATC
```

FIG.\_41H

91 / 154



**FIG.\_42A**

92 / 154

```

NcoI
~~~~~
KpnI
~
1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
 GTACCCGGTC CATATTAATA CCTATAGAG TTCGTTTATT AGCTTTATAG TGGTAACCGA TGTATATAGAC

71 AGCTCCGAGT TCTGACTGCA GTC TGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC
 TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAAAC TAGATCTTGA GATCTATCGT GTCGGTGTCTG

141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTGT GAGACGGAGC TCTTTCCCTAC CTCCTGACGT
 TGGATGTCTT CACGCTGTGA ACACCTGACA TCATCACAAC CTCTGCCCTCG AGAAAGGATG GAGGACTGCA

211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAC TCACGCGCTC CCAACAAAAT ATCGTCCCCC
 ACGCGGCCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGCGCGAG GGTGTGTTTA TAGCAGGGGG

281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTGCT GAATCTCGCT TCCACTGGCC
 TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACACA CTTAGAGCGA AGGTGACCGG

351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCT TCACCCCCTGG CGTCATGGGA
 TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCCCT

421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAAACCCAAT
 ACCTTTCTTT GGAGGCAACG AGCCTACTCA GTCGGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA

SphI
~~~~~
491  TCAGATTCCC  CCAATAGAGA  AAGTATAGCA  TGCTTTCGGG  TTTTGTGTTGG  CTTAATTGAC  TTTATTTTGTG
   AGTCTAAGGG  GGTATATCTT  TTCAATATCGT  ACGAAAGCCC  AAAACAAACC  GAATTAACTG  AAATAAAAC

561  TTGGAGTTGA  ATGCTGATTT  GTTGTGTAAA  ATGCCCAACC  ATCTGAATAT  CGAGACGGAT  AATAGGCTGG
   AACCTCAACT  TACGACTAAA  CAACACATTT  TACGGGTTGG  TAGACTTATA  GCTCTGCCTA  TTATCCGACC

```

FIG.\_42B



93 / 154

```

631  CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
    GATTAAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

    PstI
    ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
 AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCGTTTACC ACTTCGCACC GCTACTCTAC CCATATTTTG

771 CCCCCGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCCTGC TCCCCCTGCC GGACGACCCA
 GGGGCCGTGG CCTGCGCTC GAGGGCGGAT GGTCAATGGTA GAGCGGAGCG AGGGGGACGG CCTGCTGGGT

841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGMCGTGC ACAAGGAGGT SAACTTCGTS GCCTACCTCC
 CATTTATGA CAACGGGTGA GCGGCCGCTC TACCKGCACG TGTTCCTCCA STTGAAGCAS CGGATGGAGG

 NcoI
    ~~~~~
911  TGATCGTSCT CGGCCTCCTC TTGCTCGTST CCGCCATGGA GCACGTGGAC GCCAAGGCCT GCACCCKCGA
    ACTAGCASGA GCCGGAGGAG AACGAGCASA GGCGGTACCT CGTGCACCTG CGGTTCCGGA CGTGGGMGCT

    NotI
    ~~~~~
981 GTGCGGCAAC CTCGGCTTCG GCATCTGCCC GGCGGCCGCC TCCACGCAGG GCATCTCCGA AGACCTCTAC
 CACGCCGTTG GAGCCGAAGC CGTAGACGGG CCGCCGGCGG AGGTGCGTCC CGTAGAGGCT TCTGGAGATG

 SalI
    ~~~~~
    AccI
    ~~~~~
1051 AGCCGTTTAG TCGAAAATGGC CACTATCTCC CAAGCTGCCT ACGCCGACCT GTGCAACATT CCGTCGACTA
 TCGGCAAAATC AGCTTTACCG GTGATAGAGG GTTCGACGGA TCGGGCTGGA CACGTTGTAA GGCAGCTGAT

 BamHI
    ~~~~~
1121 TTATCAAGGG AGAGAAAATT TACAATTCTC AAACGTGACAT TAACGGATGG ATCCTCCGCG ACGACAGCAG
    AATAGTTCCC TCTCTTTTAA ATGTTAAGAG TTTGACTGTA ATTGCCCTACC TAGGAGGCGC TGCTGTCTGC
    
```

FIG..42C

94 / 154

```

1191 CAAAGAAATA ATCACCCTCT TCCGTGGCAC TGGTAGTGAT ACGAATCTAC AACTCGATAC TAACTACACC
      GTTTCCTTTAT TAGTGGCAGA AGGCACCGTG ACCATCACTA TGCTTAGATG TTGAGCTATG ATTGATGTGG

1261 CTCACGCCCTT TCGACACCCCT ACCACAATGC AACGGTTGTG AAGTACACCGG TGGATATTAT ATTGGATGGG
      GAGTGGGGAA AGCTGTGGGA TGGTGTACG TTGCCAACAC TTCATGTGCC ACCTATAATA TAACCTACCC

1331 TCTCCGTCCA GGACCAAGTC GAGTCGCTTG TCAAAACAGCA GGTTAGCCAG TATCCGGACT ACGCGCTGAC
      AGAGGCAGGT CCTGGTTCAG CTCAGCGAAC AGTTGTCTGT CCAATCGGTC ATAGGCCCTGA TCGCGGACTG

1401 CGTGACCGGC CACKCCCTCG GCGCCTCCCT GCGGGCACTC ACTGCCGCC AGCTGTCTGC GACATACGAC
      GCACTGGCCG GTGMGGGAGC CGCGGAGGGA CCGCCGTGAG TGACGGCGGG TCGACAGACG CTGTATGCTG

1471 AACATCCGCC TGTACACCTT CGGCGAACCG CGCAGCGGCA ATCAGGCCCTT CGCGTCGTAC ATGAACGATG
      TTGTAGGCGG ACATGTGGAA GCCGCTTGGC GCGTCGCCGT TAGTCCGGAA GCGCAGCATG TACTTGCTAC

      XhoI
      ~~~~~

1541 CCTTCCAAGC CTCGAGCCCA GATACGACGC AGTATTTCCG GGTCACTCAT GCCAACGACG GCATCCCAAA
 GGAAGGTCG GAGCTCGGGT CTATGCTGCG TCATAAAGGC CCAGTGAGTA CCGTTGCTGC CGTAGGGTTT

 NcoI
      ~~~~~

1611 CCTGCCCCCG GTGGAGCAGG GGTACGCCCA TGGCGGTGTA GAGTACTGGA GCGTTGATCC TTACAGCGCC
      GGACGGGGGC CACCTCGTCC CCATGCGGGT ACCGCCACAT CTCATGACCT CGCAACTAGG AATGTCGCGG

1681 CAGAACACAT TTGTCTGCAC TGGGGATGAA GTGCAGTGCT GTGAGGCCCA CACTCCGGGT CCCGCCCTGTC CCACACTTAT
      GTCTTGTTGTA AACAGACGTG ACCCTACTTT CACGTCACGA CACTCCGGGT CCCGCCCTGTC CCACACTTAT

1751 ATGCGCACAC GACTTATTTT GGGATGACGA GCGGAGCCTG TACATGGTGA TCAGTCATTT CAGCCTCCCC
      TACGCGTGTG CTGAATAAAA CCCTACTGCT CGCCTCGGAC ATGTACCACT AGTCAGTAAA GTCGGAGGGG

1821 GAGTGTACCA GGAAAGATGG ATGTCCTGGA GAGGGGGCCG CGTAACCACT GAAGGATGAG CTGTAAAGAA
      CTCACATGGT CCTTCTACC TACAGGACCT CTCCCCCGGC GCATTGGTGA CTTCCCTACTC GACATTTCTT

```

FIG. 42D

+

95 / 154

```
1891 GCAGATCGTT CAAACATTTG GCAATAAAGT TTCTTAAGAT TGAATCCTGT TGCCGGTCTT GCGATGATTA
CGTCTAGCAA GTTTGTAAAC CGTTATTCTA AAGAAATTCTA ACTTAGGACA ACGGCCAGAA CGCTACTAAT

1961 TCATATAAAT TCTGTTGAAT TACGTTAAGC ATGTAATAAT TAACATGTAA TGCATGACGT TATTTATGAG
AGTATATTAA AGACAACCTA ATGCAATTCTG TACATTATTA ATTGTACATT ACGTACTGCA ATAAATACTC

2031 ATGGGTTTTT ATGATTAGAG TCCCGCAATT ATACATTTAA TACGCGATAG AAAACAATAA ATAGCGCGCA
TACCCAAAAA TACTAATCTC AGGCGGTAA TATGTAAATT ATGCGCTATC TTTTGTTTTA TATCGCGCGT

                BssHII
                ~~~~~
 ClaI HindIII
                ~~~~~
                XbaI
                ~~~~~

2101 AACTAGGATA AATTATCGCG CGCGGTGTCA TCTATGTTAC TAGATCGATA AGCTTCTAGA GCGGCCGGTG
TTGATCCTAT TTAATAGCGC GCGCCACAGT AGATACAATG ATCTAGCTAT TCGAAGATCT CGCCGGCCAC

 BssHII
                ~~~~~

2171 GAGCTCCAAT TCGCCCTATA GTGAGTCGTA TTACGCGCGC TCACTGGCCG TCGTTTTACA ACGTCGTGAC
CTCGAGGTTA AGCGGGATAT CACTCAGCAT AATGCGCGCG AGTGACCGGC AGCAAAATGT TGCAGCACTG

2241 TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCG AGCATCCCC CACATGCCAGC TGGCGTAATA
ACCCTTTGG GACCGCAATG GGTGGAATTA GCGGAACGTC GTGTAGGGG AAAGCGGTG ACCGCATTAT

2311 GCGAAGAGGC CCGCACCGAT CGCCCTTCCC AACAGTTGCG CAGCCTGAAT GGCGAATGGG ACGCGCCCTG
CGCTTCTCCG GCGGTGGCTA GCGGGAAGGG TTGTCAACGC GTCGGACTTA CCGCTTACCC TCGCGGGGAC

2381 TAGCGGCGCA TTAAGCGCGG CGGGTGTGGT GGTACGCGC AGCGTGACCG CTACACTTGC CAGCGCCCCA
ATCGCCGCGT AATTCGCGC GCCCACACCA CCAATGCGCG TCGCACTGGC GATGTGAACG GTCGCGGGAT

2451 GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTTCTCGCCA CGTTCGCGCG CTTTCCCGGT CAAGCTCTAA
CGCGGCGGAG GAAAGCGAAA GAAGGGAAGG AAAGAGCGGT GCAAGCGGCC GAAAGGGGCA GTTCGAGATT
```

FIG.\_42E

+

96 / 154

2521 ATCGGGGGCT CCTTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAAC TTGATTAGGG  
TAGCCCCCGA GGGAAATCCC AAGGCTAAAT CACGAAATGC CGTGGAGCTG GGGTTTTTTG AACTAATCCC

2591 TGATGGTTCA CGTAGTGGG CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTTGGA GTCCACGTTT  
ACTACCAAGT GCATCACCCG GTAGCGGGAC TATCTGCCAA AAGCGGGA ACTGCAACCT CAGGTGCAAG

2661 TTTAATAGTG GACTCTTGT CCAAACTGGA ACAACACTCA ACCCTATCTC GGCTATATTCT TTTGATTATAT  
AAATTATCAC CTGAGAACAA GGTTTGACCT TGTGTGAGT TGGGATAGAG CCAGATAAGA AAACATAATA

2731 AAGGGATTTT GCCGATTTTG GCCTATTGGT TAAAAAATGA GCTGATTATA CAAAAATTTA ACGCCGAATTT  
TTCCCTAAAA CGGCTAAAGC CGGATAACCA ATTTTACT CGACTAAAT GTTTTAAAT TCGCGCTTAAA

2801 TAACAAAAATA TTAACGCTTA CAATTTAGGT GGCACTTTTT GGGGAAATGT GCGCGGAACC CCTATTGTGT  
ATTGTTTTAT AATTGCGAAT GTTAAATCCA CCGTGAAAAG CCCCTTTACA CGCGCCTTGG GGATAAACAA

2871 TATTTTTCTA AATACATTCA AATATGTATC CGCTCATGAG ACAATAACCC TGATAAATGC TTCAATAATA  
ATAAAAAGAT TTATGTAAGT TTATACATAG GCGAGTACTC TGTATTGGG ACTATTACG AAGTTATTAT

2941 TTGAAAAAGG AAGAGTATGA GTATTCAACA TTTCCGTGTC GCCCTTATTC CCTTTTTTGC GGCATTTTGC  
AACTTTTTCC TTCTCATACT CATAAGTTGT AAAGGCACAG CCGGAATAAG GGAAAAAACG CCGTAAAAACG

3011 CTTCCCTGTTT TTGCTCACCC AGAAACGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG  
GAAGGACAAA AACGAGTGGG TCTTTGCGAC CACTTTCATT TTCTACGACT TCTAGTCAAC CCACGTGCTC

3081 TGGGTTACAT CGAACTGGAT CTCAACAGCG GTAAGATCCT TGAGAGTTTT TGCCCCGAAG AACGTTTTTC  
ACCCAATGTA GCTTGACCCTA GAGTTGTCCG CATTCTAGGA ACTCTCAAAA GCGGGGCTTC TTGCAAAAAG

3151 AATGATGAGC ACTTTTAAAG TTCTGTATG TGGCGCGGTA TTATCCCGTA TTGACGCCGG GCAAGAGCAA  
TTACTACTCG TGAAAAATTC AAGACGATAC ACCGCGCAT AATAGGCAAT AACTGCGGCC CGTTCTCGTT

3221 CTCGGTCGCC GCATACACTA TTCTCAGAAAT GACTTGGTTG AGTACTCACC AGTCACAGAA AAGCATCTTA  
GAGCCAGCGG CGTATGTGAT AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTCGTAGAAT

3291 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT AACCATGAGT GATAACACTG CGGCCAATTT  
GCCTACCGTA CTGTCATTCT CTTAATACGT CACGACGGTA TTGGTACTCA CTATTGTGAC GCCGGTTGAA

FIG. 42F

97 / 154

```

3361 ACTTCTGACA ACGATCGGAG GACCGAAGGA GCTAACCGCT TTTTTCACACA ACATGGGGGA TCATGTAACT
    TGAAGACTGT TGCTAGCCTC CTGGCTTCCT CGATTGGCGA AAAAACGTGT TGTACCCCTT AGTACATTGA

3431 CGCCTTGATC GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAACGACGA GCGTGACACC ACGATGCCCTG
    GCGGAAC TAG CAACCTTGG CCTCGACTTA CTTCGGTATG GTTTGCTGCT CGCACTGTGG TGCTACGGAC

3501 TAGCAATGGC AACAAAGTTG CGCAAACTAT TAACTGGCGA ACTACTTACT CTAGCTTCCC GGCAACAATT
    ATCGTTACCG TTGTTGCAAC GCGTTTGATA ATTGACCGCT TGATGAATGA GATCGAAGGG CCGTTGTAA

3571 AATAGACTGG ATGGAGGCGG ATAAAGTTGC AGGACCACCTT CTGCGCTCGG CCCTTCGGC TGGCTGGTTT
    TTATCTGACC TACCTCCGCC TATTTCAACG TCCTGTGAA GACGCGAGCC GGAAGGCCG ACCGACCAA

3641 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGGTCTCGCG GTATCATTCG AGCACTGGG CCAGATGGTA
    TAACGACTAT TTAGACCTCG GCCACTCGCA CCCAGAGCGC CATAGTAACG TCGTGACCCC GGTCTACCAT

3711 AGCCCTCCCG TATCGTAGTT ATCTACACGA CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT
    TCGGGAGGC ATAGCATCAA TAGATGTGCT GCCCTCAGT CCGTTGATAC CTACTTGCTT TATCTGTCTA

3781 CGCTGAGATA GGTGCCCTCAC TGATTAAGCA TTGGTAACTG TCAGACCAAAG TTTACTCATA TATACTTTAG
    GCGACTCTAT CCACGGAGTG ACTAATTCTGTA AACCATTTGAC AGTCTGGTTC AAATGAGTAT ATATGAAATC

3851 ATTGATTTAA AACTTCATTT TTAATTTAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCTATGACCA
    TAACTAAATT TTGAAGTAAA AATTAATTT TCCTAGATCC ACTTCTAGGA AAAACTATTA GAGTACTGGT

3921 AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCGTAGAA AAGATCAAAG GATCTTCTTG
    TTTAGGGAAT TGCACCTCAA AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTCTAGTTTC CTAGAAGAAC

3991 AGATCCCTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA AAAAACCAC CGCTACCAGC GGTGGTTTGT
    TCTAGGAAAA AAAGACGCGC ATTAGACGAC GAACGTTTGT TTTTGTGGTG GCGATGGTGC CCACCAAACA

4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA
    AACGGCCTAG TTCTCGATGG TTGAGAAAAA GGCTTCCATT GACCGAAGTC GTCTCGCGTC TATGGTTTAT

4131 CTGTCTTCT AGGTAGCCG TAGTTAGGCC ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC
    GACAGGAAGA TCACATCGGC ATCAATCCGG TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG
    
```

FIG.\_42G

98 / 154

4201 TCTGCTAATC CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGA CTCAAGA  
AGACGATTAG GACAAATGGTC ACCGACGACG GTCACCGCTA TTCAGCACAG AATGGCCCCA CCTGAGTTCT

4271 CGATAGTTAC CGGATAAGGC GCAGCGGTGC GGCTGAACGG GGGGTTTCGTG CACACAGCCC AGCTTGGAGC  
GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC CCCCAAGCAC GTGTGTGCGG TCGAACCTCG

4341 GAACGACCTA CACCGAACTG AGATACCTAC AGCGTAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG  
CTTGCTGGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTTTCG CCGTGCGAAG GGCTTCCCTC

4411 AAAGCGGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA CGAGGGAGCT TCCAGGGGA  
TTTCCGCCCTG TCCATAGGCC ATTCGCCGTC CCAGCCTTGT CCTCTCGCGT GCTCCCTCGA AGTCCCCCT

4481 AACGCCTGGT ATCTTTATAG TCCTGTGCGG TTTCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT  
TTGCGGACCA TAGAAATATC AGGACAGCCC AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA

4551 CGTCAGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCCTTTTTA CCGTTCTCTG CCTTTTGTG  
GCAGTCCCC CGCCTCGGAT ACCTTTTTGC GGTCTGTTGCG CCGGAAAAAT GCCAAGGACC GGAACACGAC

4621 GCCTTTTGCT CACATGTTCT TTCCCTGCGTT ATCCCCGTAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG  
CGGAAAAACGA GTGTACAAGA AAGGACGCAA TAGGGGACTA AGACACCTAT TGGCATAATG GCGGAACTC

4691 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC  
ACTCGACTAT GCGGAGCGGC GTCGGCTTGC TGGCTCGCGT CGCTCAGTCA CTCGCTCCTT CGCCTTCTCG

4761 GCCCAATACG CAAACCGCCT CTCCTCCGCGC GTTGGCCGAT TCATTAATGC AGCTGGCAGC ACAGGTTTCC  
CGGGTTATGC GTTTGGCGGA GAGGGGCGCG CAACCGGCTA AGTAATTACG TCGACCGTGC TGTCCAAAGG

4831 CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGGC ACCCCAGGCT  
GCTGACCTTT CGCCCGTCAC TCGCCGTTGCG TTAATTACAC TCAATCGAGT GAGTAATCCG TGGGGTCCGA

4901 TTACACTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAATTTCAC ACAGGAAACA  
AATGTGAAAT ACGAAGGCCG AGCATACAAC ACACCTTAAC ACTCGCCTAT TGTAAAGTG TGTCCTTTGT

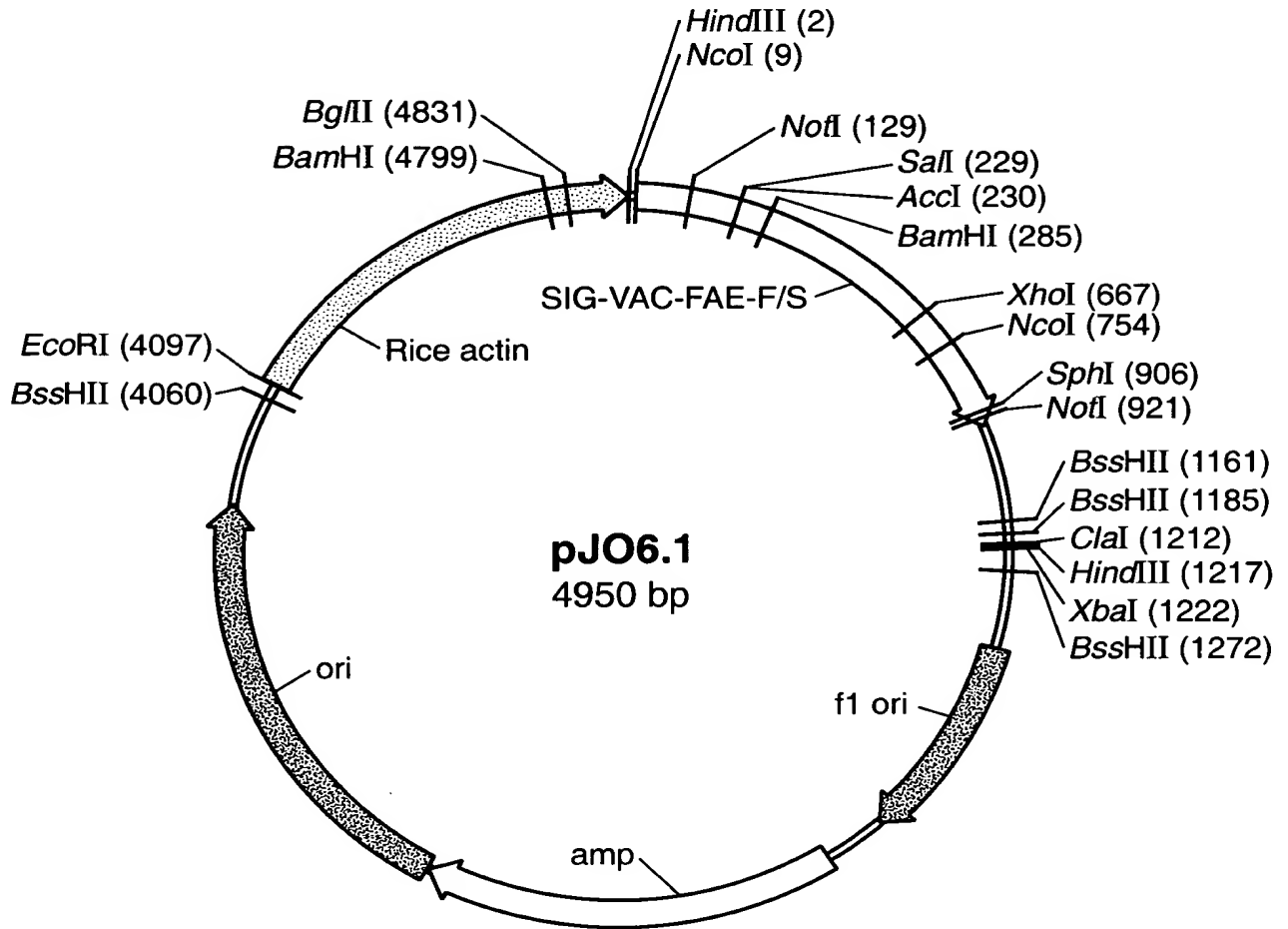
FIG.\_42H

4971 GCTATGACCA TGATTACGCC AAGCGCGCAA TTAACCCCTCA CTAAAGGGAA CAAAAGCTGG GTAC  
CGATACTGGT ACTAATGCGG TTCGCGCGTT AATTGGGAGT GATTTCCTTT GTTTTCGACC CATG

NcoI  
KpnI  
~~~~~  
BssHII
~~~~~

FIG.\_42I

100 / 154



**FIG. 43A**



HindIII NcoI  
~~~~~  
1 AAGCTTACCA TGGCCCAACG CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GCCACGGCC GCCGTCGCCG
TTCGAATGGT ACCGGGTGCG GCGCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC

NotI
~~~~~  
71 TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCGATCCG GCCCGTCACC GACCGCGGG CCGCCTCCAC  
AGCGGAGGAG GAGGAGGAAG CCGCTGAGGT TGGGCTAGGC CGGGCAGTGG CTGGCGCGCC GCGGAGGTG

SalI  
~~~~~  
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
CGTCCCCTAG AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTTC ACGGATGCGG

AccI
~~~~~  
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAAC TACATTAACG  
CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCTCTCT TTTAAATGTT AAGAGTTTGA CTGTAATTGC

BamHI  
~~~~~  
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCAC TGGTA GTGATACGAA
CTACCTAGGA GCGCTGCTG TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAATC GATACTAAT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
AGATGTTGAG CTATGATTGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG TTACGTTGCC AACACTTCAT

421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AGTCGAGTC GCTTGTCAA CAGCAGGTTA
GTGCCACCTA TAATATAACC TACCCAGAGG CAGGTCTCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT

491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
CGGTACATAG CCTGATGCGC GACTGGCACT GGCCGGTGMG GGAGCCCGCG AGGGACCGCC GTGAGTGACG

561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GCGGACATG TGAAGCCGC TTGGCGCGTC GCCGTTAGTC

FIG. 43B

102 / 154

| | | | |
|------|---|----------------|----------------|
| | XhoI | | |
| 631 | GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA
CGGAAGCGCA GCATGTACTT GCTACGGAAG GTTCGGAGCT CCGGTCTATG CTGCGTCATA AAGGCCCAGT | | |
| | ~~~~~ | | |
| 701 | CTCATGCCAA CGACGGCATC CCAAACCTGC CCCC GGGTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA
GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG CCGGTACCGC CACATCTCAT | NcoI
~~~~~ | |
| 771 | CTGGAGCGTT GATCCTTACA GCGCCCAAGAA CACATTGTC TGCACTGGGG ATGAAGTGCA GTGCTGTGAG
GACCTCGCAA CTAGGAATGT CGCGGTCTT GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC | | |
| 841 | GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGCACCT
CGGGTCCCGC CTGTCCCAACA CTATTACGC GTGTGCTGAA TAAACCCCTA CTGCTCGCCG CGTACGTGGA | SphI
~~~~~ | |
| | ~~~~~ | | |
| 911 | GGCCGGTTCG GCGCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTGGCAA
CCGGCCAGCG CCGGCGCCTT TGGTGACTTC CTACTCGACA TTTCTTCGTC TAGCAAGTTT GTAAACCGTT | | |
| 981 | TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGCGA TGATTATCAT ATAATTCTG TTGAATTACG
ATTTCAAAGA ATTCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAAGAC AACTTAATGC | | |
| 1051 | TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC
AATTCGTACA TTATTAAATG TACATTACGT ACTGCAATAA ATACTCTACC CAAAAATACT AATCTCAGGG | | |
| 1121 | GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAACT AGGATAAATT ATCGCGCGCG
CGTTAATATG TAAATTATGC GCTATCTTTT GTTTTATATC GCGCGTTTGA TCCTATTATA TAGCGCGCGC | BssHI
~~~~~ | BssHI
~~~~~ |

FIG.. 43C

103 / 154

```

XbaI
~~~~~
ClaI HindIII
~~~~~
1191 GTGTCATCTA TGTTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTTCG CCTATAGTGA
CACAGTAGAT ACAATGATCT AGCTATTCTGA AGATCTCGCC GGCACCTCG AGGTTAAGCG GGATATCACT

BssHII
~~~~~
1261 GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTTACAAAGT CGTGACTGGG AAAACCCCTGG CGTTACCCAA
CAGCATAATG CCGCGGAGTG ACCGGCAGCA AAATGTTGCA GCACTGACCC TTTTGGGACC GCAATGGGTT

1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC
GAATTAGCGG AACGTCGTGT AGGGGAAAG CCGTCGACCG CATTATCGCT TCTCCGGGCG TGGCTAGCGG

1401 CTTCCCAACA GTTGCGCAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGCGCATTA GCGCGGCGGG
GAAAGGTTGT CAACGCGTCG GACTTACCGC TTACCCCTGCG CCGGACATCG CCGCGTAATT CGCGCCGCCC

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTTCTTC
ACACCACCAA TCGCGGTCGC ACTGGCGATG TGAACGGTCG CCGGATCGCG GCGGAGGAAA GCGAAAGAAG

1541 CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCCCTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC
GGAAGGAAAG AGCGGTGCAA GCGGCCGAAA GGGGCAGTTC GAGATTTAGC CCCCAGGGA ATCCCAAGG

1611 GATTAGTGC TTTACGGCAC CTCGACCCCA AAAAAGTTGA TTAGGGTGAT GGTTCACGTA GTGGGCCATC
CTAAATCACG AAATGCCGTG GAGCTGGGGT TTTTGTGAAT ATCCCACTA CCAAGTGCAT CACCCGGTAG

1681 GCCCTGATAG ACGGTTTTC GCCCTTTGAC GTTGGAGTCC ACGTTCTTTA ATAGTGGACT CTTGTTCCAA
CGGGAATATC TGCCAAAAG CCGGAAACTG CAACCTCAGG TGCAAGAAAT TATCACCTGA GAACAAGGTT

1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATCTTTTG ATTTATAAGG GATTTGCGG ATTTCGGCCT
TGACCTTGTT GTGAGTTGGG ATAGAGCCAG ATAAGAAAAC TAAATATTCC CTAAAACGGC TAAAGCCGGA

1821 ATTGGTTAAA AAATGAGCTG ATTTAACAAA AATTAAACGC GAATTTTAAAC AAAATATTAA CGCTTACAAT
TAAACCAATT TTTACTCGAC TAAATTGTTT TTAAATTTGCG CTTAAATTTG TTTTATAAAT GCGAATGTTA

```

FIG._43D

104 / 154

```

1891  TTAGGTGGCA  CTTTTCGGGG  AAATGTGCGC  GGAACCCCTA  TTTGTTTATT  TTTCTAAATA  CATTCAAATA
      AATCCACCGT  GAAAAGCCCC  TTTACACGCG  CCTTGGGGAT  AAACAAATAA  AAAGATTTAT  GTAAGTTTAT

1961  TGTATCCGCT  CATGAGACAA  TAACCCCTGAT  AAATGCTTCA  ATAATATTGA  AAAAGGAAGA  GTATGAGTAT
      ACATAGGCCG  GTACTCTGTT  ATTGGGACTA  TTTACGAAGT  TATTATAACT  TTTTCCTTCT  CATACTCATA

2031  TCAACATTTC  CGTGTGCCCC  TTATTCCCTT  TTTTGGGGCA  TTTTGCCTTC  CTGTTTTTGC  TCACCCAGAA
      AGTTGTAAAG  GCACAGCGGG  AATAAGGGA  AAAACGCCGT  AAAACGGAAG  GACAAAACG  AGTGGGTCTT

2101  ACGCTGGTGA  AAGTAAAAGA  TGCTGAAGAT  CAGTTGGGTG  CACGAGTGGG  TTACATCGAA  CTGGATCTCA
      TGCAGACCACT  TTCATTTTCT  ACGACTTCTA  GTCAACCCAC  GTGCTCACCC  AATGTAGCTT  GACCTAGAGT

2171  ACAGCGGTAA  GATCCTTGAG  AGTTTTCGCC  CCGAAGAACG  TTTTCCAATG  ATGAGCACTT  TTAAGTTTCT
      TGTGCGCCATT  CTAGGAACTC  TCAAAAGCGG  GGCCTCTTGC  AAAAGTTAC  TACTCGTGAA  AATTCAAGA

2241  GCTATGTGGC  GCGGTATTAT  CCCGTATTGA  CGCCGGGGCA  GAGCAACTCG  GTCGCCGCAT  ACACATTCTT
      CGATACACCG  CGCCATAATA  GGGCATAACT  CGGCCCCGTT  CTCGTTGAGC  CAGCGGCGTA  TGTGATAAGA

2311  CAGAATGACT  TGGTTGAGTA  CTCACCAGTC  ACAGAAAAGC  ATCTTACGGA  TGGCATGACA  GTAAGAGAAAT
      GTCTTACTGA  ACCAACTCAT  GAGTGGTCAG  TGTCTTTTCG  TAGAATGCCCT  ACCGTACTGT  CATTCCTCTTA

2381  TATGCAGTGC  TGCCATAACC  ATGAGTGATA  ACACTGCGGC  CAACCTACTT  CTGACAACGA  TCGGAGGACC
      ATACGTCACG  ACGGTATTGG  TACTCACTAT  TGTGACGCCG  GTTGAATGAA  GACTGTTGCT  AGCCTCCTGG

2451  GAAGGAGCTA  ACCGCTTTT  TGCACAACAT  GGGGATCAT  GTAACTCGCC  TTGATCGTTG  GGAACCGGAG
      CTTCTCTCGAT  TGGCGAAAAA  ACGTGTGTGA  CCCCCTAGTA  CATTGAGCGG  AACTAGCAAC  CCTTGGCCTC

2521  CTGAATGAAG  CCATACCAAA  CGACGAGCGT  GACACCACGA  TGCCTGTAGC  AATGGCAACA  ACGTTGCGCA
      GACTTACTTC  GGTATGGTTT  GCTGTCTCGA  CTGTGGTGCT  ACGGACATCG  TTACCCTTGT  TGCAACGCGT

2591  AACTATTAA  TGGCGAACTA  CTTACTCTAG  CTTCCCGGCA  ACAATTAAATA  GACTGGATGG  AGGCGGATAA
      TTGATAATTG  ACCGCTTGAT  GAATGAGATC  GAAGGGCCGT  TGTTAATTAT  CTGACCTACC  TCCGCCCTATT

2661  AGTTGCAGGA  CCACTTCTGC  GCTCGGCCCT  TCCGGCTGGC  TGGTTTATTG  CTGATAAATC  TGGAGCCGGT
      TCAACGTCCT  GGTGAAGACG  CGAGCCGGGA  AGCCCGACCG  ACCAAATAAC  GACTATTTAG  ACCTCGGCCA
    
```

FIG._43E

105 / 154

2731 GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGGCCAG ATGGTAAGCC CTCCCGTATC GTAGTTATCT
CTCGCACCCA GAGCGCCATA GTAACGTCGT GACCCCGGTC TACCATTCGG GAGGCGATAG CATCAATAGA

2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT
TGCTGTGCCC CTCAGTCCGT TGATACCTAC TTGCTTTATC TGCTTAGCGA CTCTATCCAC GGAGTGAATA

2871 TAAGCATTTG TAACTGTCAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAACT TCATTTTAA
ATTTCGTAACC ATTGACAGTC TGGTTCAAAT GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAAAATT

2941 TTTAAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CCTTAAACGT GAGTTTTCGT
AAATTTTCCT AGATCCACTT CTAGGAAAAA CTATTAGAGT ACTGGTTTA GGAATTGCA CTCAAAAAGCA

3011 TCCACTGAGC GTCAGACCCC GTAGAAAAAG TCAAAAGGATC TTCTTGAGAT CCTTTTTC TGCGCGTAAT
AGGTGACTCG CAGTCTGGGG CATCTTTTCT AGTTCCCTAG AAGAACTCTA GGAATAAAG ACGCGCATTA

3081 CTGCTGCTTG CAAACAAAAA AACCAACCGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT
GACGACGAAC GTTTGTTTTT TTGGTGCGGA TGGTCGCCAC CAAACAAACG CCTAGTTCT CGATGGTTGA

3151 CTTTTTCCGA AGGTAACCTG CTTCAGCAGA GCGCAGATAC CAAATACTGT CCTTCTAGTG TAGCCGTTAGT
GAAAAAGGCT TCCATTGACC GAAGTCGTCT CGCGTCTATG GTTTATGACA GGAAGATCAC ATCGGCATCA

3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC
ATCCGGTGGT GAAGTTCTTG AGACATCGTG GCGGATGTAT GGAGCGAGAC GATTAGGACA ATGCTCACCG

3291 TGCTGCCAGT GCGGATAAGT CGTGCTTTAC CGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGCGCAG
ACGACGGTCA CCGCTATTCA GCACAGAAATG GCCCAACCTG AGTTCTGCTA TCAATGGCCT ATTCCGCGTC

3361 CGGTCGGGCT GAACGGGGG TTCTGTGCACA CAGCCCAGCT TGGAGCGAAC GACCTACACC GAACGTGAGT
GCCAGCCCCA CTTGCCCCCC AAGCACGTGT GTCGGGTCTG ACCTCGCTTG CTGGATGTGG CTTGACTCTA

3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG
TGGATGTCGC ACTCGATACT CTTTCGCGGT CGGAAGGGCT TCCCCTTTC CGCCTGTCCA TAGGCCATTC

3501 CGGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGGGAAACG CCTGGTATCT TTATAGTCCT
GCCGTCCTC CTTGTCTC TCGCGTGCTC CCTCGAAGGT CCCCCTTTC GGACCATAGA AATATCAGGA

FIG._43F

106 / 154

```

3571  GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA
      CAGCCCAAAG CGGTGGAGAC TGAACCTCGA GCTAAAAACA CTACGAGCAG TCCCCCGGCC TCGGATACCT

3641  AAAACGCCAG CAACGCGGCC TTTTACGGT TCCTGGCCTT TTGCTGGCCT TTTGCTCACA TGTCTTTTCC
      TTTTGGCGTC GTTGGCGCGG AAAAAATGCCA AGGACCGGAA AACGACCGGA AACGAGTGT ACAAGAAAGG

3711  TGCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCAGC
      ACGCAATAGG GGAATAAGAC ACCTATTGGC ATAATGCGG AACTCACTC GACTATGGCG AGCGGCGTCG

3781  CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGGCC AATACGCAA CCGCCTCTCC
      GCTTGCTGGC TCGCGTCGCT CAGTCACTCG CTCCTTCGCC TTCTCGCGGG TTATGCGTTT GCGGAGAGG

3851  CCGCGCGTTG GCCGATTCTA TAATGCAGCT GGCACGACAG GTTTCCTCGAC TGGAAAGCGG GCAGTGAGCG
      GCGCGCAAC CGGCTAAGTA ATTACGTCGA CCGTGTGTC CAAAGGCTG ACCTTTCGCC CGTCACTCGC

3921  CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
      GTTGCCTTAA TTACACTCAA TCGAGTGAGT AATCCGTGGG GTCCGAAATG TGAATACGA AGGCCGAGCA

      BssHII
      ~~~~

3991  ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC
      TACAACACAC CTTAACACTC GCCTATTGTT AAAGTGTGTC CTTTGTCTGAT ACTGGTACTA ATGCGGTTTCG

      BssHII
      ~~~~
      EcoRI
      ~~~~~~

4061  GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGAATT CCACAATGAA CAATAATAAG ATTAAAAATAG
      CGCGTTAATT GGGAGTGATT TCCCTTGTTT TCGACCTTAA GGTGTTACTT GTTATTATTC TAATTTTATC

4131  CTTGCCCCCG TTGCAGCGAT GGGTATTTT TCTAGTAAAA TAAAAGATAA ACTTAGACTC AAAACATTTA
      GAACGGGGGC AACGTCGCTA CCCATAAAA AGATCATTTT ATTTTCTATT TGAATCTGAG TTTTGTAAAT

4201  CAAAAACAAC CCCTAAAGTC CTAAAGCCCA AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA
      GTTTTGTGTG GGGATTTCAG GATTTCGGGT TTCACGATAC GTGCTAGGTA TCGTTCGGGT CGGGTTGGGT
  
```

FIG._43G

107 / 154

```

4271 ACCCAACCCA ACCCAACCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTGAGT
    TGGGTGGGT TGGGTGGGT CACGTGGGT GACCGTTTAT CAGAGGTGGG GCGCGTGATA GTGGCACTCA

4341 TGTCCGCACC ACCGCACGTC TCGCAGCCAA AAAAAAAAA AGAAGAAAA AAAAACAAGC
    ACAGGCGTGG TGGCGTGCAG AGCGTCGGTT TTTTTTTTTT TCCTTCTTTT TTTTCTTTT CTTTTGTGCG

4411 AGGTGGGTCC GGTCTGTGGG GGCCGGAAAA GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC
    TCCACCCAGG CCCAGACCC CCGGCTTTT CGCTCCTCCT AGCGTCTGTC GCTGCTCCGG GCCGGGAGGG

4481 TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATATACA TACCCCCCCC TCCTCTCCCA TCCCCCAAC
    AGCGAAGGT TTCTTTGCGG GGGTAGCGG TGATATATGT ATGGGGGGG AGAGGAGGGT AGGGGGGTG

4551 CCTACCACCA CCACCAACCAC CACCTCCTCC CCCCTCGCTG CCGGACGACG AGCTCCTCCC CCCTCCCCCT
    GGATGGTGGT GGTGGTGGTG GCCATTGGTG GGGCGGGGAG AGGAGAAAGA AAGAGGCAAA AAAAAAGCA GAGCCAGAGC

4621 CCGCCGCCGC CGGTAACCAC CCCGCCCTC TCCTCTTTCT TCCTCCTGTT TTTTTCGT CTCGGTCTCG
    GCGCGCGGCG GCCATTGGTG GGGCGGGGAG GGGGAGCGAC GCGCTGCTGC TCGAGGAGGG GGGAGGGGGA

4691 ATCTTTGGCC TTGGTAGTTT GGGTGGCGA GAGCGGCTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG
    TAGAAACCGG AACCATCAA CCCACCCGCT CTCGCCAAG CAGCGGTCT AGCCACGCGC CCTCCCCGCC

                                BamHI                                BglII
                                ~~~~~                                ~
4761 GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA
    CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCT AGGAGCGCCC CTTACCCCGA GAGCCTACAT

                                BglII
                                ~~~~~
4831 GATCTCTTTT CTTTCTTCTT TTTGTGGTAG AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTCCTT
    CTAGAAGAAA GAAAGAAGAA AACACCATC TTAAACTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAGAA

4901 TTCATGATTT GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGGTAG
    AAGTACTAAA CACTGTTTAC GTCGGAGCAC GCCTCGAAAA AACATCCATC
  
```

FIG..43H

108 / 154

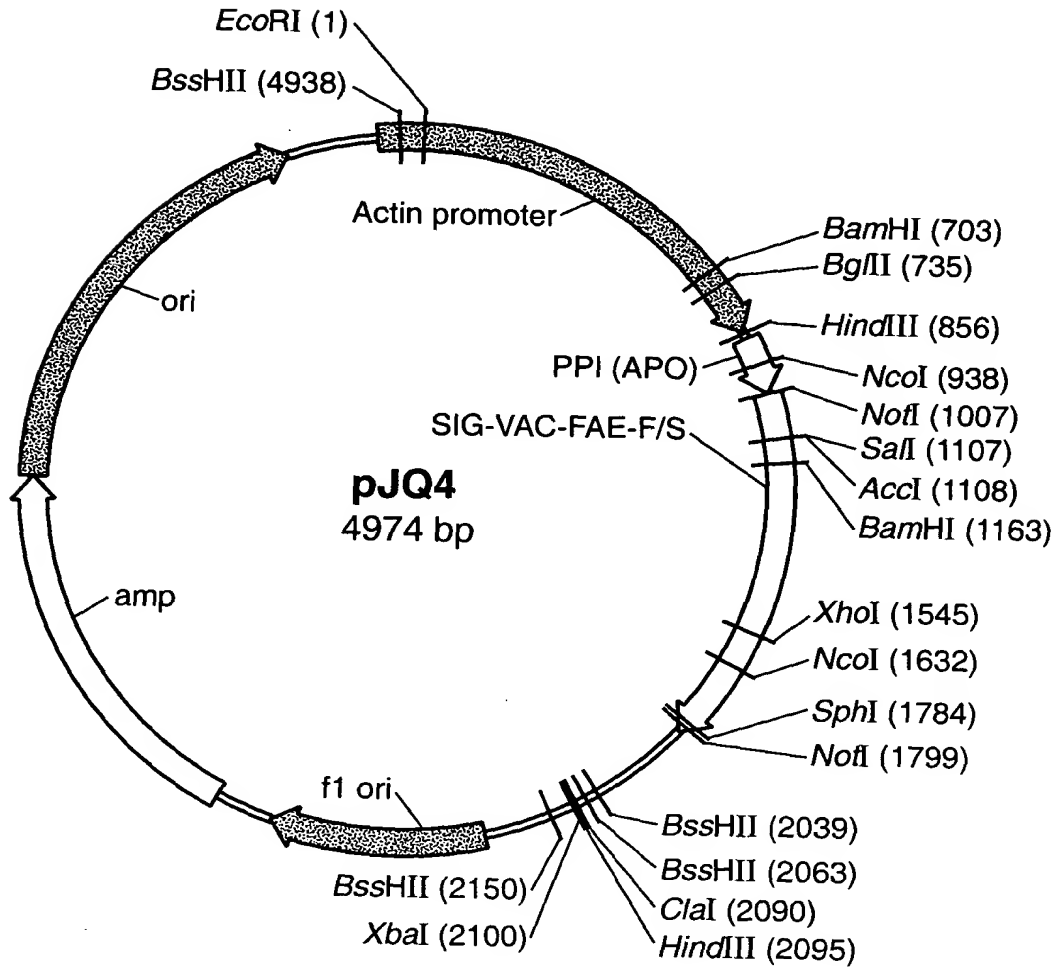


FIG. 44A

ECORI
~~~~~  
1 AATTCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTTAGT  
TTAAGGTGTT ACTTGTTATT ATCTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA  
  
71 AAAATAAAG ATAAACTTAG ACTCAAAAACA TTTACAAAAA CAACCCCTAA AGTCCTAAAG CCCAAAGTGC  
TTTTATTTC TATTGAATC TGAGTTTTGT AAATGTTTTT GTTGGGATT TCAGGATTTC GGGTTTCACG  
  
141 TATGCACGAT CCATAGCAAG CCCAGCCCAA CCCAACCAC CCCAGTGCAG CCAACTGGCA  
ATACGTGCTA GGTATCGTTC GGGTCGGGT GGGTTGGGT GGGTCACGTC GGTGACCGT  
  
211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAA  
TTATCAGAGG TGGGGGCCGT GATAGTGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTT  
  
281 AAAAAGAAAG AAAAAGAAAG AAAAAGAAAG CAGCAGGTGG GTCCGGGTGG TGGGGGCCGG AAAAGCGAGG  
TTTTTCTTC TTTTTTTTCT TTTTCTTTT GTCTCCACC CAGGCCCAGC ACCCCCGGCC TTTTCGCTCC  
  
351 AGGATCGCGA GCAGCGACGA GGCCCGGCC TCCCTCCGCT TCCAAAGAAA CGCCCCCCAT CGCCACTATA  
TCCTAGCGCT CGTCGCTGCT CCGGGCCGG AGGAGGCCA AGTTTCTTT GCGGGGGTA GCGGTGATAT  
  
421 TACATACCCC CCCCTCTCCT CCCATCCCC CCACCTACC ACCACCACCA CCACCACCTC CTCCCCCTC  
ATGTATGGG GGGGAGAGGA GGTAGGGG GTTGGATGG TGGTGGTGG GGTGGTGGAG GAGGGGGAG  
  
491 GCTGCCGGAC GACGAGCTCC TCCCCCTCC CCCCTCCGG CGCCCGGTAA CCACCCCGCC CCTCTCCTCT  
CGACGGCCTG CTGCTCGAG AGGGGGGAG GGGAGCGGC GCGGGCCATT GGTGGGGCGG GGAGAGGAGA  
  
561 TTCTTTCTCC GTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCTTTGGTA GTTTGGGTGG GCGAGAGCGG  
AAGAAAGAGG CAAAAAAGAG AGCAGAGCCA GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC  
  
631 CTTCGTCGCC CAGATCGGTG CCGGGGAGGG GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC  
GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC GCGCCCTAGAG CCGCGACCGC AGAGGCCCGC ACTCAGCCCG  
  
BamHI  
~~~~~  
701 CGGATCCCTCG CGGGGAATGG GGCCTCTCGA TGATAGTCTT CTTCCTTTCT TCTTTTGTG GTAGAATTG
GCCTAGGAGC GCCCCTTACC CCGAGAGCCT ACATCTAGAA GAAAGAAAGC AGAAAAACAC CATCTTAAAC

FIG.. 44B

110 / 154

```
771  AATCCCTCAG CATTGTTTCAT CGGTAGTTTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT AATGCAGCCT CGTGCGGAGC
    TAGGGGAGTC GTAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA TTACGTCGGA GCACGCCCTCG

    HindIII
    ~~~~~
841  TTTTTTGTAG GTAGAAGCTT ACNATGCMCG TGCACAAGGA GGTSAACTTC GTSGCCTACC TCCTGATCGT
    AAAAAACATC CATCTTCGAA TGKTACCKGC ACGTGTTCCT CCASTTGAAG CASC GGATGG AGGACTAGCA

    NcoI
    ~~~~~
911  SCTCGGCTC CTCTTGCTCG TSTCCGCCAT GGAGCACGTG GACGCCAAGG CCTGCACCCCK CGAGTGCGGGC
    SGAGCCGGAG GAGAACGAGC ASAGGCGGTA CCTCGTGCAC CTGCGGTTCC GGACGTGGGM GCTCACGCCCG

    NotI
    ~~~~~
981  AACCTCGGCT TCGGCATCTG CCCGGCGGCC GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT
    TTGGAGCCGA AGCCGTAGAC GGGCCGCCGG CGGAGGTGCG TCCC GTAGAG GCTTCTGGAG ATGTCGGCAA

    SalI
    ~~~~~
    AccI
    ~~~~~
1051 TAGTCGAAAT GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA
    ATCAGCTTTA CCGGTGATAG AGGTTTCGAC GGATGCGGCT GGACACGTTG TAAGGCAGCT GATAATAGTT

    BamHI
    ~~~~~
1121 GGGAGAGAAA ATTTACAATT CTCAAACTGA CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA
    CCCTCTCTTT TAAATGTTAA GAGTTTGAAT GTAATTGCCT ACCTAGGAGG CGCTGCTGTC GTCGTTTCTT

1191 ATAATCACCG TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTCGA TACTAACTAC ACCCTCACGC
    TATTAGTGGC AGAAGGCACC GTGACCATCA CTATGCTTAG ATGTTGAGCT ATGATTGATG TGGAGTGCG

1261 CTTTTCGACAC CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGTCTTCCGT
    GAAAGCTGTG GGATGGTGTT ACGTTGCCAA CACTTCATGT GCCACCTATA ATATAACCTA CCCAGAGGCA
```

FIG. 44C

111 / 154

```

1331 CCAGGACCAA GTCGAGTCGC TTGTCAAACA GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC
      GGTCCCTGGTT CAGCTCAGCG AACAGTTTGT CGTCCAATCG GTCATAGGCC TGATGCGCGA CTGGCACTGG

1401 GGCCACKCCC TCGGCGCCTC CCTGGCGGCA CTCACCTGCCG CCCAGCTGTC TCGACATAC GACAACATCC
      CCGGTGMGG AGCCGCGGAG GGACCGCCGT GAGTGACGGC GGTTCGACAG ACGCTGTATG CTGTTGTAGG

1471 GCCTGTACAC CTTCGGCGAA CCGCGCAGCG GCAATCAGGC CTTCCGCTCG TACATGAACG ATGCCCTTCCA
      CGGACATGTG GAAGCCGCTT GCGCGCTCGC CGTTAGTCCG GAAGCGCAGC ATGTACTTGC TACGGAAGGT

      XhoI
      ~~~~~
1541 AGCCTCGAGC CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC
      TCGGAGCTCG GGTCTATGCT GCGTCATAAA GGCCCAGTGA GTACGGTTGC TGCCGTAGGG TTTGGACGGG

      NcoI
      ~~~~~
1611 CCGGTGGAGC AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACA
      GGCCACCTCG TCCCCATGCG GGTACCGCCA CATCTCATGA CCTCGCAACT AGGAATGTCG CGGGTCTTGT

1681 CATTTGTCTG CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA
      GTAAACAGAC GTGACCCCTA CTTACAGTCA CGACACTCCG GGTCCCGCCT GTCCACACACT TATTACGCGT

      SphI
      ~~~~~
      NotI
      ~~~~~
1751 CACGACTTAT TTTGGGATGA CGAGCGGCGC ATGCACCTGG CCGGTCGCGG CCGCGGAAC CACTGAAGGA
      GTGCTGAATA AAACCTACT GCTCGCCGCG TACGTGGACC GGCCAGCGCC GGCGCCTTG GTGACTTCCT

1821 TGAGCTGTAA AGAAGCAGAT CGTTCAAACA TTTGGCAATA AAGTTTCTTA AGATTGAATC CTGTTGCCGG
      ACTCGACATT TCTTCGTCTA GCAAGTTTGT AAACCGTTAT TTCAAAGAAT TCTAACTTAG GACAACGGCC

1891 TCTTGCATG ATTATCATAT AATTCTGTT GAATTACGTT AAGCATGTAA TAATTAACAT GTAATGCATG
      AGAACGCTAC TAATAGTATA TTAAGACAA CTTAATGCAA TTCTGTACATT ATTAATTGTA CATTACGTAC

1961 ACGTTATTTA TGAGATGGGT TTTTATGATT AGAGTCCCGC AATTATACAT TTAATACGCG ATAGAAAACA
      TGCRAATAAT ACTCTACCCA AAAATACTAA TCTCAGGGCG TTAATATGTA AATTATGCGC TATCTTTTGT

```

FIG._44D

112 / 154

```

XbaI
~~~~~
          ClaI  HindIII
          ~~~~~
2031  AAATATAGCG CGCAAACTAG GATAAATTAT CGCGCGCGGT GTCACTCTATG TTACTAGATC GATAAGCTTC
      TTTATATCGC GCGTTTGATC CTATTTAATA GCGCGCGCCA CAGTAGATAC AATGATCTAG CTATTCGAAG

          BssHII
          ~~~~~
          BssHII
          ~~~~~
XbaI
~~~~~
2101  TAGAGCGGCC GGTGGAGCTC CAAATCGCCC TATAGTGAGT CGTATTACGC GCGCTCACTG GCCGTCGTTT
      ATCTCGCCGG CCACCTCGAG GTTAAGCGGG ATATCACTCA GCATAATGCG CGCGAGTGAC CGGCAGCAAA

2171  TACAACGTCG TGACTGGGAA AACCCCTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC CCCCTTTTCG
      ATGTTGCAGC ACTGACCCCTT TTGGGACCGC AATGGTTGA ATTAGCGGAA CGTCGTGTAG GGGGAAAGCG

2241  CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT TCGCAGCCCT GAATGGCGAA
      GTCGACCGCA TTATCGCTTC TCCGGGCGTG GCTAGCGGGA AGGTTGTCA ACGGTCGGA CTTACCGCTT

2311  TGGGACGCGC CCTGTAGCGG CGCATTAAGC GCGCGGGTG TGGTGGTTAC GCGCAGCGTG ACCGCTACAC
      ACCCTGCGCG GGACATCGCC GCGTAATTCTG CGCGGCCAC ACCACCAATG CGCGTCGCAC TGGCGATGTG

2381  TTGCCAGCGC CCTAGCGCCC GCTCCTTTCTG CTTTCTTCCC TTCCTTTCTC GCCACGTTCTG CCGGCTTTCC
      AACGGTCGCG GGATCGCGGG CGAGGAAAGC GAAAGAGGG AAGGAAAGAG CCGTGCAAGC GGCCGAAAGG

2451  CCGTCAAGCT CTAAATCGGG GGCTCCCTTT AGGTTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAAA
      GGCAGTTCTGA GATTAGCCC CCGAGGGAAA TCCCAAGGCT AATCAACGAA ATGCCGTGGA GCTGGGGTTT

2521  AAACCTGATT AGGGTGATGG TTCACGTAGT GGGCCATCGC CCTGATAGAC GGTTTTTCGC CCTTTGACGT
      TTTGAACTAA TCCCACTACC AAGTGCAATCA CCCGGTAGCG GGACTATCTG CCAAAAAGCG GGAAACTGCA

2591  TGGAGTCCAC GTTCTTTAAT AGTGGACTCT TGTTCCAAAC TGGAACAACA CTCAAACCCTA TCTCGGTCTA
      ACCTCAGGTG CAAGAAATTA TCACCTGAGA ACAAGGTTTG ACCTTGTTGT GAGTTGGGAT AGAGCCAGAT

2661  TTCTTTTGAT TTATAAGGA TTTTGCCGAT TTCGGCCCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAAAA
      AAGAAAACTA AATATCCCT AAAACGGCTA AAGCCGGATA ACCAATTTT TACTCGACTA AATTGTTTTT

```

FIG.-44E

113 / 154

```
2731 TTTAACGCGA ATTTTAACAA AATATTAACG CTTACAATTT AGGTGGCACT TTTCGGGGAA ATGTGCGCGG
AAATTGCGCT TAAAAATTGTT TTATAATTGC GAATGTTAAA TCCACCCGTA AAAGCCCCCTT TACACGCGCC

2801 AACCCCTATT TGTTTATTTT TCTAAATACA TTCAAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA
TTGGGGATAA ACAAATAAAA AGATTATGT AAGTTTATAC ATAGCGAGT ACTCTGTTAT TGGGACTATT

2871 ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCG TGTCGCCCTT ATTCCCTTTT
TACGAAGTTA TTATAACTTT TTCTTCTCTCA TACTCATAAG TTGTAAAGC ACAGCGGAA TAAGGGAATA

2941 TTGCGGCATT TTGCCCTTCCT GTTTTGTCTC ACCCAGAAAC GCTGGTGAAG GTAAAAAGATG CTGAAGATCA
AACGCCGTAA AACGGAAGGA CAAAAACGAG TGGGTCTTTG CGACCACTTT CATTTCTAC GACTTCTAGT

3011 GTTGGGTGCA CGAGTGGGT ACATCGAACT GGATCTCAAC AGCGGTGAAG TCCTTGAGAG TTTTCGCCCC
CAACCCACGT GCTCACCCAA TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAACCTCTC AAAAGCGGGG

3081 GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCG GGTATATATCC CGTATTGACG
CTTCTTGCAA AAGTTACTA CTCGTGAAAA TTTCAAGACG ATACACCGCG CCATAATAGG GCATAACTGC

3151 CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC
GGCCCGTTCT CGTTGAGCCA GCGGCGTATG TGATAAGAGT CTTACTGAAC CAACTCATGA GTGGTCAGTG

3221 AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA TGCAGTGCTG CCATAACCAT GAGTGATAAC
TCTTTTCGTA GAATGCCCTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGGTA CTCACTATTG

3291 ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG
TGACGCCGGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT TCCTCGATTG GCGAAAAAAC GTGTTGTACC

3361 GGGATCATGT AACTCGCCTT GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA
CCCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA CTTACTTCGG TATGGTTTGC TGCTCGCACT

3431 CACCACGATG CCTGTAGCAA TGGCAACAACT GTTGGCGCAA CTATTAACTG GCGAACTACT TACTCTAGCT
GTGGTGCTAC GGACATCGTT ACCGTTGTTG CAACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA

3501 TCCCGGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC
AGGCCCGTTG TTAATTATCT GACCTACCTC GCCTATTTC ACGTCCCTGG TGAAGACGCG AGCCGGGAAG
```

FIG._44F

114 / 154

3571 CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT
GCCGACCGAC CAAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT AACGTCGTGA
3641 GGGGCCAGAT GGTAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA
CCCCGGTCTA CCATTGCGGA GGGCATAGCA TCAATAGATG TGCTGCCCTT CAGTCCGTG ATACCTACTT
3711 CGAAATAGAC AGATCGCTGA GATAGGTGCC TCACTGATTA AGCATTTGGTA ACTGTCAGAC CAAGTTTACT
GCTTTATCTG TCTAGCGACT CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAAATGA
3781 CATATATACT TTAGATTGAT TTAAAACTTC ATTTTAAAT TAAAGGATC TAGGTGAAGA TCCTTTTTGA
GTATATATGA AATCTAACTA AATTTTGAAG TAAAAATTAA ATTTTCTTAG ATCCACTTCT AGGAAAAACT
3851 TAATCTCATG ACCAAAAATCC CTTAACGTGA GTTTTCGTTT CACTGAGCGT CAGACCCCGT AGAAAAAGATC
ATTAGAGTAC TGGTTTTAGG GAATTGCACT CAAAAGCAAG GTGACTCGCA GTCTGGGCA TCTTTTCTAG
3921 AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAAA CCACCGCTAC
TTTCCTAGAA GAACCTCTAGG AAAAAAGAC GCGCATAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG
3991 CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC
GTCGCCACCA AACAAACGGC CTAGTTCTCG ATGGTTGAGA AAAAGGCTTC CATTGACCGA AGTCGTCTCG
4061 GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCCACCACCT TCAAGAATCT TGTAGCACCG
CGTCTATGGT TTATGACAGG AAGATCACAT CGGCATCAAT CCGGTGGTGA AGTTCTTGAG ACATCGTGGC
4131 CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACC
GGATGTATGG AGCGAGACGA TTAGGACAAAT GGTCAACCGAC GACGGTCACC GCTATTTCAGC ACAGAATGGC
4201 GGTTGGACTC AAGACGATAG TTACCGGATA AGCGCGAGCG GTCGGGCTGA ACGGGGGT CCGTGCACACA
CCAACTGAG TTCTGCTATC AATGGCCTAT TCCGCGTCCG CAGCCCGACT TGCCCCCAA GCACGTGTGT
4271 GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG
CGGGTCGAAC CTCGCTTGCT GGATGTGGCT TGACTCTATG GATGTCGCAC TCGATACTCT TTCGCGGTGC
4341 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG
GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTCCG CGTCCCAGCC TTGTCCTCTC GCGTGTCTCC

FIG._44G

115 / 154

4411 AGCTTCCAGG GGAACGCG TGGTATCTTT ATAGTCTCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCTG
TCGAAGGTCC CCTTTGCGG ACCATAGAAA TATCAGGACA GCCCCAAGCG GTGGAGACTG AACTCGCAGC

4481 ATTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGA AACCAGCA ACAGGCGCTT TTTACGGTTC
TAAAAACACT ACAGCAGTC CCCCCGCTC GGATACCTTT TTGCGGTCTG TCGCCCGGAA AATGCCAAG

4551 CTGGCCCTTT GCTGGCCTTT TGCTCACATG TTCTTTCTCTG CGTTATCCCC TGATTCTGTG GATAACCGTA
GACCGGAAA CGACCGGAAA ACGAGTGATC AAGAAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT

4621 TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGACGCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA
AATGGCGGAA ACTCACTCGA CTATGGCGAG CGGCGTCGGC TTGCTGGCTC GCGTCGCTCA GTCACCTCGCT

4691 GGAAGCGGAA GAGCGCCCAA TACGCAAAAC CCTCTCTCCC GCGCGTTGGC CGATTCTATTA ATGCAGCTGG
CCTTCGCCCTT CTCGCGGGTT ATGCGTTTGG CGGAGAGGGG CGCGCAACCG GCTAAGTAAT TACGTCGACC

4761 CACGACAGGT TTCCCGACTG GAAAGCGGCG AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACCTCAT
GTGCTGTCCA AAGGCTGAC CTTTCGCCCG TCACTCGCGT TGCGTTAATT AACTCAATC GAGTGAGTAA

4831 AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT
TCCGTGGGGT CCGAAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA

BssHII
~~~~~

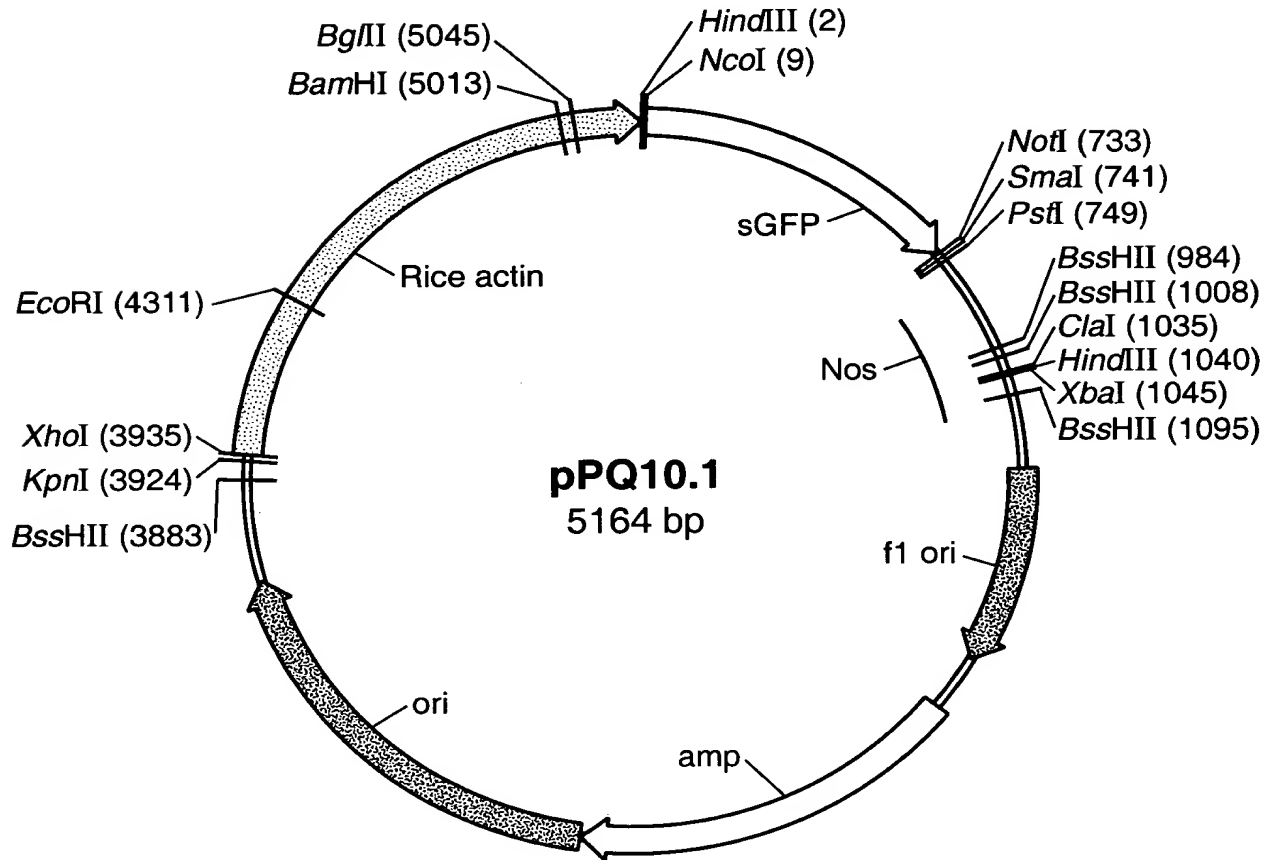
4901 TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCGC GCAATTAAAC CTCACTAAAG GGAACAAAAG  
AGTGTGTCCT TTGTCGATAC TGGTACTAAT GCGGTTCCG CGTTAATTGG GAGTGATTTT CCTTGTGTTT

EcoR

4971 CTGG  
GACC

FIG.\_44H

116 / 154



**FIG. 45A**



117 / 154

```

HindIII NcoI
~~~~~
1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCG GGTGGTGCC CATCTGGTC GAGCTGGACG
 TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACCACGG GTAGGACCAG CTCGACCTGC

71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT GCCACCTACG GCAAGCTGAC
 CGCTGCACTT GCCGGTGTTC AAGTCGCACA GGCCTGCTCC GCTCCCGCTA CGGTGGATGC CGTTCGACTG

141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCCCACCC TCGTGACCAC CTTCACCTAC
 GGACTTCAAG TAGACGTGGT GGCCTGTTCTGA CCGGCACGGG ACCGGGTGGG AGCACTGGTG GAAGTGGATG

211 GCGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGCACTT CTTCAAGTCC GCCATGCCCCG
 CCGCACGTCA CGAAGTCGGC GATGGGCTG GTGTACTTCG TCGTGTGAA GAAGTTCAGG CCGTACGGGC

281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA
 TTCCGATGCA GGTCCCTCGC TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT

351 GTTCGAGGCG GACACCCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC
 CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTTGTAG

421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAG CAGAAGAACG
 GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGCTGTTC GTCTTCTTGC

491 GCATCAAGGT GAACCTCAAG ATCCGCCACA ACATCGAGGA CCGCAGCGTG CAGCTCGCCG ACCACTACCA
 CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGATGGT

561 GCAGAACACC CCCATCGGCG ACGGCCCCGT GCTGCTGCCC GACAACCACT ACCTGAGCAC CCAGTCCGCC
 CGTCTTGTGG GGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA TGGACTCGTG GGTACGGCGG

631 CTGAGCAAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCGGGGATCA
 GACTCGTTTC TGGGGTTGCT CTTCCGCGTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

```

FIG.\_45B

**FIG. 45C**

119 / 154

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGGGC GGGTGTGGTG GTTACGGCA GCGTGACCGC TACACTTGCC  
GCGCGGACA TCGCCGCGTA ATTCGCGCG CCCACACAC CAATGCGCGT CGCACTGGCG ATGTGAACGG

1331 AGCGCCCTAG CGCCCGCTCC TTTTCGTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGC TTTCCCCGTC  
TCGCGGATC GCGGCGGAG AAGCGAAG AAGGGAAG AAGAGCGGTG CAAGCGGCCG AAAGGGGCAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTAGGGT TCCGATTAG TGCTTTACGG CACCTCGACC CCAAAAACT  
TTCGAGATT AGCCCCCGAG GAAATCCCA AGGCTAAATC ACGAAATGCC GTGGAGCTGG GGTTTTGA

1471 TGATTAGGT GATGGTTCAC GTAGTGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG  
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGACT ATCTGCCAAA AGCGGGAAA CTGCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTGTTC CAAACTGGAA CAACACTCAA CCCTATCTCG GTCTATTCTT  
AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT GGGATAGAGC CAGATAAGAA

1611 TTGATTATA AGGATTTTG CCGATTTCCG CCTATTGGT AAAAAATGAG CTGATTAAAC AAAAATTAA  
AACTAAATAT TCCCTAAAC GCTAAAGCC GGATAACCAA TTTTACTC GACTAAATG TTTTAAAT

1681 CGCGAATTTT AACAAAATAT TAACGCTTAC AATTAGGTG GCACTTTTCG GGGAAATGTG CGCGGAACCC  
GCGCTTAAAA TTGTTTATA ATTGCGAATG TTAATCCAC CGTGAAAAGC CCTTTACAC GCGCCTTGGG

1751 CTATTTGTTT ATTTTCTAA ATACATTCAA ATATGTATCC GTCATGAGA CAATAACCCCT GATAAATGCT  
GATAAACAAA TAAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTACGA

1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT TTCCCGTGTG CCCTTATTCC CTTTTTTGCG  
AGTTATTATA ACTTTTCTCT TCTCATACTC ATAAGTTGTA AAGGCACAGC GGAATAAGG GAAAAAACGC

1891 GCATTTTGCC TTCTTGTTT TGCTCACCCA GAAACGCTGG TGAAGTAA AGATGCTGAA GATCAGTTGG  
CGTAAACCG AAGGACAAA ACGAGTGGGT CTTTGCAGC ACTTTCATTT TCTACGACTT CTAGTCAACC

1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTTT GCCCCGAAGA  
CACGTGCTCA CCCAATGTAG CTTGACCCTAG AGTTGTCGCC ATCTAGGAA CTCTCAAAAG CCGGGCTTCT

2031 ACGTTTCCA ATGATGAGCA CTTTTAAAGT TCTGCTATGT GCGCGGGTAT TATCCCGTAT TGACGCCGGG  
TGCAAAAGGT TACTACTCGT GAAAATTCA AGACGATACA CCGCGCCATA ATAGGCATA ACTGCGGCC

FIG. 45D

120 / 154

2101 CAAGAGCAAC TCGGTCGCCG CATACTAT TCTCAGAATG ACTTGGTTGA GTACTACCA GTACACAGAAA  
GTTCTCGTTG AGCCAGCGC GTATGTGATA AGAGTCTTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC  
TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG

2241 GGCCAACTTA CTTCAGCAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGGAT  
CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGCGAA AAAACGTGTT GTACCCCTTA

2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA  
GTACATTGAG CGGAACCTAGC AACCTTGGC CTCGACTTAC TTCGGTATGG TTTGCTGCTC GCACTGTGGT

2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAAACCTAT AACTGGCGAA CTACTTACTC TAGCTTCCC  
GCTACGGACA TCGTTACCGT TGTGCAACG CGTTGATAA TTGACCGCTT GATGAATGAG ATCGAAGGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGCCTCGGC CCTTCCGGCT  
CGTTGTTAAT TATCTGACCT ACCTCCGCCCT ATTTCAACGT CCTGGTGAAG ACGCGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCG TATCATTGCA GCAC TGGGG  
CCGACCAAAT AACGACTATT TAGACCTCGC CCACTCGCAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA  
GTCTACCAAT CCGGAGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCCCTCACT GATTAAGCAT TGGTAACTGT CAGACCAAGT TTAATCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTATAA ACTTCATTTT TAATTTAAAA GGATCTAGGT GAAGATCCTT TTTGATAATC  
TATGAAATCT AACTAAATTT TGAAGTAAAA ATTAATTTT CCTAGATCCA CTTCTAGGAA AAATAATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG  
AGTACTGGTT TTAGGGAATT GCACCTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCTTTT TTCTGCGCGT AATCTGCTGC TTGCAACAA AAAAACCAAC GCTACCAAGC  
TAGAAGAACT CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGT TTTTGGTGG CGATGTCGC

FIG.\_45E

121 / 154

2941 GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA  
CACCAAAACA ACGGCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAAGTCG TCTCGCGTCT

3011 TACCAAAATAC TGTCCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG CACCGCCTAC  
ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG  
TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCCAA

3151 GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCCG GCTGAACGGG GGGTTCGTGC ACACAGCCCCA  
CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTGCGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC  
CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT  
GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTC CTCTCGCGTG CTCCCCTCGAA

3361 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT  
GGTCCCCCTT TCGGGACCAT AGAAATATCA GGACAGCCCCA AAGCGGTGGA GACTGAAC TC GCAGCTAAAA

3431 TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGA AAAACGC CAGCAACGCG GCCTTTTAC GGTTCCTGGC  
AACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG GTCGTTGCGC CGGAAAAATG CCAAGGACCG

3501 CTTTGTGCTG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCTTGATT CTGTGGATAA CCGTATTACC  
GAAAACGACC GGAAACGAG TGTACAAGAA AGGACGCAAT AGGGACTAA GACACCTATT GGCATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG  
CGGAAACTCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGCGTC GCTCAGTCAC TCGCTCCCTTC

3641 CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCGCGCGG TTGGCCGATT CATTAATGCA GCTGGCACGA  
GCCTTCTCGC GGGTTATGCG TTTGGCGGAG AGGGCGCGC AACCGGCTAA GTAATTACGT CGACCGTGCT

3711 CAGGTTTCCC GACTGGAAG CGGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTAGGCA  
GTCCAAAGGG CTGACCTTTC GCCCGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT

FIG.\_45F

122 / 154

```
3781 CCCACGAGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTCACA
 GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCCATT GTTAAAGTGT

 BssHII
      ~~~~~
3851  CAGGAAACAG  CTATGACCAT  GATTACGCCA  AGCGGCAAT  TAACCCTCAC  TAAAGGGAAC  AAAAGCTGGG
      GTCCTTTGTC  GATACTGGTA  CTAATGCGGT  TCGCGCGTTA  ATTGGGAGTG  ATTTCCCTTG  TTTTCGACCC

      KpnI
      ~~~~~
3921 TACCGGGCCC CCCCTCGAGG TCATTTCATAT GCTTGAGAAG AGAGTCGGGA TAGTCCAAAA TAAACAAAAG
 ATGGCCCGGG GGGGAGCTCC AGTAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTTT ATTTTGTTTC

3991 GTAAGATTAC CTGGTCAAAA GTGAAACACAT CAGTTAAAAG GTGGTATAAG TAAAATATCG GTAATAAAAG
 CATCTAATG GACCAGTTTT CACTTTTGTG GTCAATTTTC CACCATATTC ATTTATATAG CATTATTTTC

4061 GTGGCCCCAA GTGAAATTTA CTCCTTTTCTA CTATTATAAA AATTGAGGAT GTTTTGTCGG TACTTTTGATA
 CACCGGGTTT CACTTTAAAT GAGAAAAGAT GATAATATTT TTAACCTCCTA CAAAACAGCC ATGAAACTAT

4131 CGTCATTTT GTATGAATTG GTTTTAAAGT TTATTCGCGA TTTGGAAATG CATATCTGTA TTTGAGTCGG
 GCAGTAAAA CATACTTAAC CAAAAATTCA AATAAGCGCT AAACCTTTAC GTATAGACAT AAACTCAGCC

4201 TTTTAAAGTT CGTTGCTTTT GTAAATACAG AGGGAATTGT ATAAGAAATA TCCTTAAAAA ACCCATATGC
 AAAAAATCAA GCAACGAAAA CATTTATGTC TCCCTAAACA TATTCCTTAT AGAAATTTT TGGGTATACG

 EcoRI
      ~~~~~
4271  TAATTGACA  TAATTTTGA  GAAAAATATA  TATTCAGGCG  AATCCACAA  TGAACAATAA  TAAGATTAAA
      ATTAAACTGT  ATTAAAAACT  CTTTTTATAT  ATAAGTCCGC  TTAAGGTGTT  ACTTGTATT  ATCTAAATT

4341  ATAGCTTGC  CCCGTTGCAG  CGATGGGTAT  TTTTCTAGT  AAAATAAAAG  ATAAACTTAG  ACTCAAAACA
      TATCGAACGG  GGGCAACGTC  GCTACCCATA  AAAAGATCA  TTTTATTTTC  TATTTGAATC  TGAGTTTGT

4411  TTTACAAAA  CAACCCCTAA  AGTCCCTAAAG  CCCAAAGTGC  TATGCACGAT  CCATAGCAAG  CCCAGCCCCA
      AAATGTTTT  GTTGGGGATT  TCAGGATTC  GGGTTTCAG  ATACGTGCTA  GGTATCGTTC  GGTCTGGGTT
```

FIG.\_45G

123 / 154

```

4481 CCCAACCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA AATAGTCTCC ACCCCCGGCA CTATCACCCGT
GGTTGGGTT GGGTTGGGTG GGGTCACGTC GGTGACCGT TTATCAGAGG TGGGGGCCGT GATAGTGGCA

4551 GAGTTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAA AAAAAGAAAG AAAAAAAGA AAAAGAAAAA
CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTTTT TTTTCTTTTC TTTTCTTTCT TTTTCTTTT

4621 CAGCAGGTGG GTCCGGGTG TGGGGGCCG AAAAGCGAGG AGGATCGGA GCAGCGACGA GGCCCGGGCC
GTCGTCCACC CAGGCCAGC ACCCCCGGCC TTTTCGCTCC TCCTAGCGT CGTCGCTGCT CCGGGCCGGG

4691 TCCCCTCCGT TCCAAGAAA CGCCCCCAT CGCCACTATA TACATACCC CCCCTCTCCT CCCATCCCCC
AGGAGGCCA AGGTTTCTTT GCGGGGGTA GCGGTGATAT ATGTATGGG GGGAGAGGA GGTAGGGGG

4761 CAACCTTACC ACCACCACCA CCACCACCTC CTCCCCCTC GCTGCCGGAC GACGAGCTC TCCCCCTCTC
GTTGGGATGG TGGTGGTGGT GGTGGTGGAG GAGGGGGAG CGACGGCCTG CTGCTCGAGG AGGGGGGAGG

4831 CCCTCCGGC CGGCCGGTAA CCACCCCGCC CCTCTCCTCT TTTCTTTCTC GTTTTTTTT TCGTCTCGGT
GGGAGGCGGC GCGGCCATT GGTGGGGCGG GGAGAGGAGA AAGAAAGAGG CAAAAAAGA AGCAGAGCCA

4901 CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG CTTCTGTCGC CAGATCGGTG CGCGGGAGGG
GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC

                               BglII
                               ~~~~~
4971 GCGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC CGGATCCTCG CGGGGAATGG GGCTCTCGGA
CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCCG GCCTAGGAGC GCCCCTTACC CCGAGAGCCT

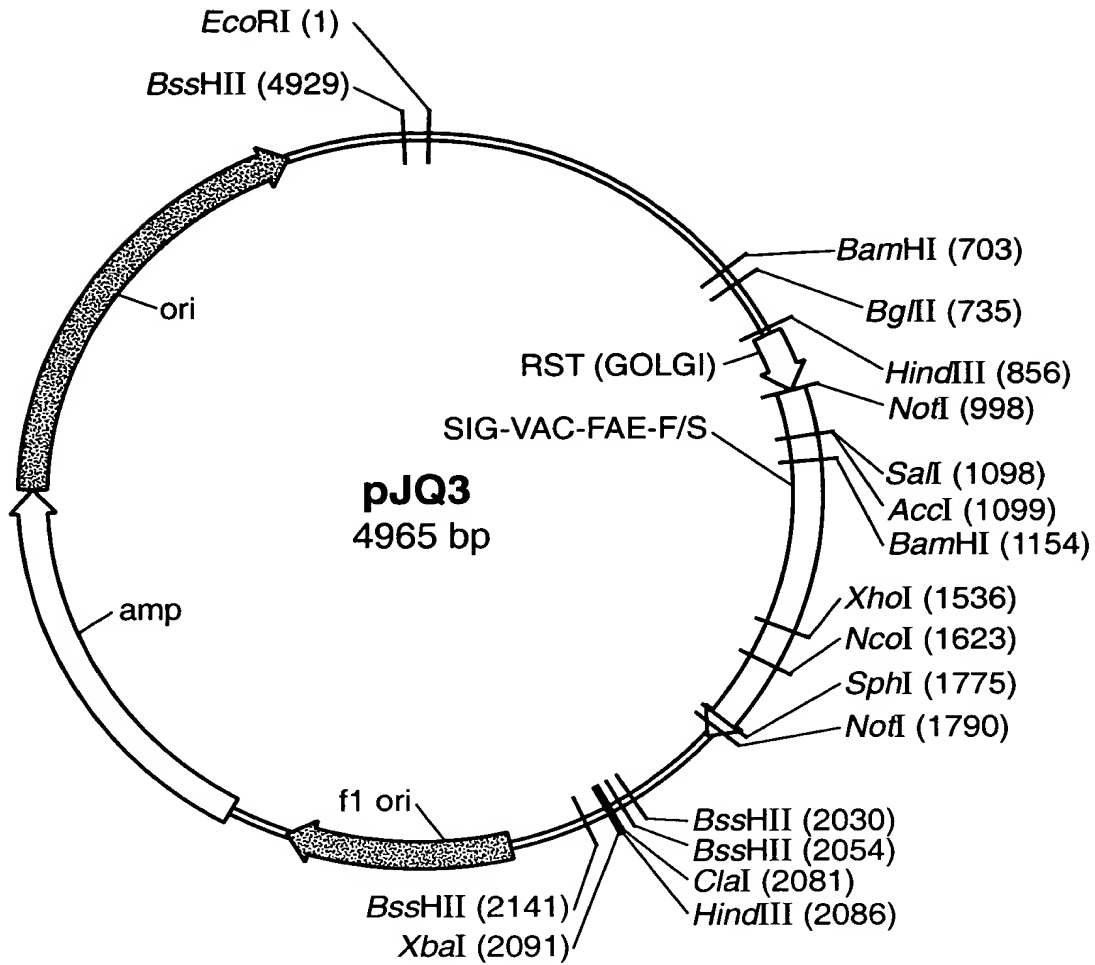
 BglII
                               ~~~~~
5041 TGTAGATCTT CTTTCTTTCT TCTTTTGTG GTAGAATTG AATCCCTCAG CATTTGTCAT CGGTAGTTTT
ACATCTAGAA GAAAGAAAGA AGAAAAACAC CATCTTAAAC TTAGGAGTC GTAACAAGTA GCCATCAAAA

5111 TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTGCGGAGC TTTTGTGTAG GTAG
AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTCG AAAAAACATC CATC

```

FIG.\_45H

124 / 154



**FIG. 46A**



125 / 154

```

EcRI
~~~~
1 AATCCACAA TGAACAATA TAAGATTAA ATAGCTTCC CCGTTGCAG CGATGGGTAT TTTTCTAGT
 TTAAGGTGTT ACTTGTTATT ATTCTAATT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA

71 AAAATAAAG ATAAACTTAG ACTCAAAACA TTTACAAAA CAACCCCTAA AGTCTTAAAG CCCAAAGTGC
 TTTTATTTTC TATTGAATC TGAGTTTGT AAATGTTTT GTTGGGATT TCAGGATTC GGGTTTCACG

141 TATGCACGAT CCATAGCAAG CCCAGCCCAA CCAACCCAC CCAGTGCAG CCAACTGGCA
 ATACGTGCTA GGTATCGTTC GGTCTGGGT GGTGTTGGTG GGTTCACGTC GGTGACCGT

211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCC CACCACCGCA CGTCTCGAG CCAAAAAAAA
 TTATCAGAGG TGGGGCCGT GATAGTGCA CTCAACAGC GTGTGGCGT GCAGAGCGTC GGTTTT

281 AAAAAGAAAG AAAAAAAGA AAAAGAAAA CAGCAGGTG GTCCGGGTG TGGGGCCGG AAAAGCGAGG
 TTTTCTTTC TTTTCTTCT TTTTCTTCT TTTTCTTCT GTCTCCACC CAGGCCAGC ACCCCCGGC TTTTCGCTCC

351 AGGATCGCA GCAGCGACA GGGCCGGCC TCCCTCCGT TCCAAAGAA CGCCCCCAT CGCCACTATA
 TCCTAGCGCT CGTCGTGCT CCGGGCCGG AGGAGGCGA AGTTTCTTT GCGGGGGTA GCGTGATAT

421 TACATACCC CCCCTCTCT CCCATCCCC CCAACCTACC ACCACACCA CCACCACTC CTCCCCCTC
 ATGTATGGG GGGAGAGGA GGTAGGGG GGTGGATGG TGTGGTGGT GGTGGTGGAG GAGGGGGAG

491 GCTGCCGGAC GACGAGCTC TCCCCCTCC CCCTCCGGC CGCCCGGTAA CCACCCCGC CCTCTCTCT
 CGACGGCCTG CTGCTCGAG AGGGGGAGG GGGAGCGGC GCGGCCATT GGTGGGGCG GGAGAGGAGA

561 TTCTTTCTCC GTTTTTTTT TCGTCTCGT TCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG
 AAGAAAGAG CAAAAAAGA AGCAGAGCA GAGCTAGAA CCGGAACCAT CAAACCCACC CGCTCTCGCC

631 CTTCGTCCG CAGATCGGT GCGGGGAGG GCGGATCTC GCGCTGGCG TCTCCGGCG TGAGTCGGC
 GAAGCAGCG GTCTAGCCAC GCGCCCTCC CGCCCTAGAG CGCCGACCG AGAGCCCGC ACTCAGCCG

 BamHI BglII
      ~~~~~                               ~~~~~
701 CGGATCCTCG CGGGGAATGG GGCTCTCGA TGATAGTCTT CTTTCTTCT TCTTTTGTG GTAGAATTG
   GCCTAGGAGC GCCCCTTACC CCGAGAGCCT ACATCTAGAA GAAAGAAAG AGAAAAACAC CATCTTAAAC

```

FIG. 46B

126 / 154

```
771 AATCCCCTCAG CATTTGTTTCAT CGGTAGTTTTT TCTTTTCATG ATTTGTGACA AATGCAGCCTT CGTGCGGAGC
TTAGGGAGTC GTAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTCG

HindIII
~~~~~
841 TTTTGTGTAG GTAGAAGCTT ACCATGATCC ACACCAACCT CAAAAAGAAG TTCTCCCTCTT TCATCCTCGT
AAAAACATC CATCTTCGAA TGGTACTAGG TGTGGTTGGA GTTTTCTTTC AAGAGGGAGA AGTAGGAGCA

911 CTTCCTCCTC TTCGCCGTGA TCTGCGTGTG GAAGAAGGGC TCCGACTACG AGGCCCTCAC CCTCCAAGCC
GAAGGAGGAG AAGCGGCACT AGACGCACAC CTTCTTCCCG AGGCTGATGC TCCGGGAGTG GGAGGTTCCG

NotI
~~~~~
981 AAGGAGTTCC AAATGGCGGC CGCCTCCACG CAGGGCATCT CCGAAGACCT CTACAGCCGT TTAGTCGAAA
TTCCCTCAAGG TTTACC GCCG GCGGAGGTGC GTCCCGTAGA GGCTTCTGGA GATGTCGGCA AATCAGCTTT

SalI
~~~~~
AccI
~~~~~
1051 TGGCCACTAT CTCCCCAAGCT GCCTACGCCG ACCTGTGCAA CATTCGTCG ACTATATATCA AGGGAGAGAA
ACCGGTGATA GAGGTTTCCA CCGATGCGGC TGGACACGTT GTAAGGCAGC TGATAATAGT TCCCTCTCTT

BamHI
~~~~~
1121 AATTACAAAT TCTCAAACTG ACATTAAACGG ATGGATCCTC CGCGACGACA GCAGCAAAGA AATAATCACC
TTAAATGTTA AGAGTTTGAC TGTAATTGCC TACCTAGGAG GCGCTGCTGT CGTCGTTTCT TTATTAGTGG

1191 GTCTTCCGTG GCACTGGTAG TGATACGAAT CTACAACTCG ATACTAACTA CACCCCTCACG CCTTTCGACA
CAGAAGGCAC CGTGACCATC ACTATGCTTA GATGTTGAGC TATGATTGAT GTGGGAGTGC GAAAAGCTGT

1261 CCCTACCACA ATGCAACGGT TGTGAAGTAC ACGGTGGATA TTATATTGGA TGGGTCTCCG TCCAGGACCA
GGGATGGTGT TACGTTGCCA ACACTTTCATG TGCCACCTAT AATATAACCT ACCCAGAGGC AGGTCCTGTT
```

FIG.- 46C

127 / 154

```
1331 AGTCGAGTCG CTTGTCAAAC AGCAGGTTAG CCAGTATCCG GACTACGCGC TGACCGGTGAC CGGCCACKCC
 TCAGCTCAGC GAACAGTTTG TCGTCCAATC GGTCAATAGGC CTGATGCGCG ACTGGCACTG GCCGGTGMGG

1401 CTCGGCGCCT CCCTGGCGGC ACTCACTGCC GCCAGCTGT CTGCGACATA CGACAAACATC CGCCTGTACA
 GAGCCGCGGA GGGACCGCCG TGAATGACGG CGGGTCGACA GACGCTGTAT GCTGTTGTAG GCGGACATGT

 XhoI
                                ~~~~~
1471 CCTTCGGCGA ACCGCGCAGC GGCAATCAGG CCTTCGCGTC GTACATGAAC GATGCCTTCC AAGCCTCGAG
    GGAAGCCGCT TGGCGCGTCG CCGTTAGTCC GGAAGCGCAG CATGTACTTG CTACGGAAGG TTCGGAGCTC

1541 CCCAGATACG ACGCAGTATT TCCGGGTAC TCATGCCAAC GACGGCATCC CAAACCTGCC CCCGGTGGAG
    GGGTCTATGC TCGGTCATAA AGGCCAGTG AGTACGGTTG CTGCCGTAGG GTTTGGACGG GGGCCACCTC

                                NcoI
                                ~~~~~
1611 CAGGGGTACG CCCATGGCGG TGTAGAGTAC TGGAGCGTTG ATCCTTACAG CGCCCAGAAC ACATTTGTCT
 GTCCCCATGC GGTACCGCC ACATCTCATG ACCTCGCAAC TAGGAATGTC GCGGGTCTTG TGTAAACAGA

1681 GCAC TGGGA TGAAGTGCAG TGCTGTGAGG CCCAGGGCGG ACAGGGTGTG AATAATGCCG ACACGACTTA
 CGTGACCCCT ACTTCACGTC ACGACACTCC GGGTCCCGCC TGTCCCACAC TTATTACGCG TGTGCTGAAT

 SphI
                                ~~~~~
                                NotI
                                ~~~~~
1751 TTTTGGGATG ACGAGCGCG CATGCACCTG GCCGGTCGCG GCCCGGAAA CCACTGAAGG ATGAGCTGTA
 AAAACCTAC TGCTCGCCGC GTACGTGGAC CGCCACGCG CGCGCCCTTT GGTGACTTCC TACTCGACAT

1821 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGCGGAT
 TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA GGACAAACGGC CAGAACGCTA

1891 GATTATCATA TAATTCTGT TGAATTACGT TAAGCATGTA ATAATTAAAC TGTAAATGCAT GACGTTATTT
 CTAATAGTAT ATTAAAGACA ACTTAATGCA ATTGCTACAT TATTAATTGT ACATTACGTA CTGCAATAAA
```

FIG.-46D

```

1961 ATGAGATGGG TTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC AAAATATAGC
 TACTCTACCC AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG TTTTATATCG
 BssHII
      ~~~
      XbaI
      ~~~~~

2031 GCGCAAACTA GGATAAATA TCGCGCGCGG TGTCACTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC
 CGCGTTTGAT CCTATTAAAT AGCGCGCGCC ACAGTAGATA CAATGATCTA GCTATTGAA GATCTCGCCG
 BssHII
      ~~~
      ClaI HindIII
      ~~~~~

2101 CGGTGGAGCT CCAATTTCGCC CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT TTACAACGTC
 GCCACCTCGA GGTAAAGCGG GATATCACTC AGCATAATGC GCGGAGTGA CCGGCAGCAA AATGTTGCAG

2171 GTGACTGGGA AAACCCCTGGC GTTACCCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTTCG CCAGCTGGCG
 CACTGACCCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA GGGGAAAAGC GGTCGACCCG
 BssHII
      ~~~

2241  TAAATAGCGA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG
      ATTATCGCTT CTCCGGGGCGT GGCTAGCGGG AAGGTTGTC AACGCGTCGG ACTTACCGCT TACCCCTGCGC

2311  CCCTGTAGCG GCGCATTAAG CGCGGCGGGT GTGGTGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG
      GGGACATCGC CGCGTAATTC GCGCCGCCCA CACCACCAAT GCGCGTCGCA CTGGCGATGT GAACGGTCCG

2381  CCCTAGCGCC CGCTCCTTTC GCTTCTTCC CTTCCTTTCT CGCCACGTTT GCCGGCTTTC CCCGTCAAGC
      GGGATCGCGG GCGAGGAAAG CGAAAGAAGG GAAGGAAAGA GCGGTGCAAG CGGCCGAAAG GGGCAGTTCG

2451  TCCTAAATCGG GGGCTCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGAT
      AGATTTAGCC CCCGAGGGAA ATCCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGGGTT TTTTGAACATA

2521  TAGGGTGATG GTTCACTAGT TGGGCCATCG CCTGATAGA CGGTTTTTTCG CCCTTTGACG TTGGAGTCCA
      ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAAGC GGGAAACTGC AACCTCAGGT
```

FIG.\_46E

129 / 154

2591 CGTCTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTCTTTTGA  
GCAAGAAATT ATCACCTGAG AACAAAGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACT

2661 TTTATAAGGG ATTTTGCCGA TTTCCGCCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG  
AAATATTCCC TAAAACGGCT AAAGCCGGAT AACCAATTTT TTACTCGACT AAATTGTTTT TAAATTGCGC

2731 AATTTTAAACA AAATATTAAC GCTTACAATT TAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT  
TTAAAAATTGT TTTATAAATG CGAATGTTAA ATCCACCGTG AAAAGCCCCCT TTACACGCGC CTTGGGGATA

2801 TTGTTTATTT TTCTAAATAC ATTCAAAATAT GTATCCGCTC ATGAGACAAAT AACCTGATA AATGCTTCAA  
AACAAATAAA AAGATTATG TAAGTTTATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGGAAGTT

2871 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTC GTGTCGCCCT TATTCCTTTT TTTGCGGCAT  
ATTATAACTT TTTCCCTTCT ATACTCATAA GTTGTAAGG CACAGCGGGA ATAAGGGA AAACGCCGTA

2941 TTTGCCCTTC TGTTTITGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAAGATC AGTTGGGTGC  
AAACGGAAGG ACAAACGA GTGGGCTTT GCGACCACTT TCATTTTCTA CGACTTCTAG TCAACCCACG

3011 ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT  
TGCTCACCCA ATGTAGCTTG ACCTAGAGTT GTCGCCATTC TAGGAACTCT CAAAAGCGGG GCTTCTTGCA

3081 TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG  
AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC GCCATAATAG GGCATAACTG CGGCCCGTTC

3151 AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCAGTCA CAGAAAAAGCA  
TCGTTGAGCC AGCGCGTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTCGT

3221 TCTTACGGAT GGCATGACAG TAAGAGAAAT ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC  
AGAATGCCTA CCGTACTGTC ATTCTCTTAA TACGTCACGA CCGTATTGGT ACTCACTATT GTGACGCGCG

3291 AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG  
TTGAATGAAG ACTGTTGCTA GCCTCCTGGC TTCTCTCGATT GGCGAAAAAA CGTGTGTGAC CCCCTAGTAC

3361 TAACTCGCCT TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT  
ATTGAGCGGA ACTAGCAACC CTTGGCCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC TGTGGTGCTA

FIG.\_46F

130 / 154

3431 GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACT GCGAACTAC TTACTCTAGC TTCCCCGGCAA  
CGGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA CCGCTTGATG AATGAGATCG AAGGGCCGTT

3501 CAATTAATAG ACTGGATGGA GCGGATATAA GTTGACAGGAC CACTTCTGCG CTCGGCCCCCTT CCGGCTGGCT  
GTTAATTATC TGACCTACCT CCGCCTATTT CAACGTCTTG GTGAAGACGC GAGCCGGGAA GCGCGACCGA

3571 GGTTTATTGC TGATAAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA  
CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG TAACGTCTGT ACCCCGGTCT

3641 TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA  
ACCAATTCGGG AGGCATAGC ATCAATAGAT GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCTTTATCT

3711 CAGATCGCTG AGATAGGTGC CTCACCTGATT AAGCATTGGT AACGTGTCAGA CCAAGTTTAC TCATATATAC  
GTCTAGCGAC TCTATCCACG GAGTGACTAA TTCGTAACCA TTGACAGTCT GGTTCAAAATG AGTATATATG

3781 TTTAGATTGA TTTAAAACTT CATTTTAAAT TTTAAAGGAT CTAGGTGAAG ATCCTTTTGG ATAATCTCAT  
AAATCTAACT AAATTTGAA GTAAAAATTA AATTTCCTA GATCCACTTC TAGGAAAAAC TATTAGAGTA

3851 GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAAGAT CAAAGGATCT  
CTGGTTTTAG GGAATTGCAC TCAAAAGCAA GGTGACTCGC AGTCTGGGC ATCTTTTCTA GTTTCCTAGA

3921 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG  
AGAACTCTAG GAAAAAAGA CGCGCATTAG ACGACGAACG TTTGTTTTTT TGGTGGCGAT GGTGCGCCACC

3991 TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTTCGAA GGTAACCTGGC TTCAGCAGAG CGCAGATACC  
AAACAAACGG CCTAGTTCTC GATGGTTGAG AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG

4061 AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC GCCTACATAC  
TTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTCTTGA GACATCGTGG CGGATGTATG

4131 CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT  
GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCAC CGCTATTGAG CACAGAAATGG CCCAACCTGA

4201 CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTGCGGCTG AACGGGGGGT TCGTGCACAC AGCCCAGCTT  
GTTCTGCTAT CAATGGCCTA TTCCGCGTCTG CCAGCCCGAC TTGCCCCCCA AGCAGGTGTG TCGGGTCCGA

FIG.\_46G

131 / 154

```

4271 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA
    CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTGCGA CTCGATACTC TTTCGCGGTG CGAAGGGCTT

4341 GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTGC GAACAGGAGA GCGCACGAGG GAGCTTCCAG
    CCTCTTTTCC GCCTGTCCAT AGGCCATTCC CCGTCCCAGC CTTGTCTCTCT CGCGTGTCTCC CTCGAAGGTC

4411 GGGGAAACGC CTGGTATCTT TATAGTCTCTG TCGGGTTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG
    CCCCTTTGCG GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACCTCGAG CTAAAAACAC

4481 ATGCTCGTCA GGGGGGCGGA GCCATATGAA AACGCCACG AACGCGGCCT TTTTACGGTT CCTGGCCCTTT
    TACGAGCAGT CCCCCCGCCT CGGATACCTT TTTGCGGTGC TTGCGCCGGA AAAATGCCAA GGACCGGAAA

4551 TGCTGGCCCTT TTGCTCACAT GTTCTTTTCTT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT
    ACGACCGGAA AACGAGTGTA CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA

4621 TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAAGCGGA
    AACTCACTCG ACTATGGCGA GCGGCGTCCG CTTGCTGGCT CCGCTCGCTC AGTCACTCGC TCCTTGCGCT

4691 AGAGCGCCCA ATACGCCAAC CGCCTCTCCC CGCGCGTTGG CCGATTCAAT AATGCAGCTG GCACGACAGG
    TCTCGCGGGT TATGCGTTTG GCGGAGAGGG GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGTGTCC

4761 TTTCCCGACT GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC
    AAAGGCTGA CCTTTCGCCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA ATCCGTGGGG

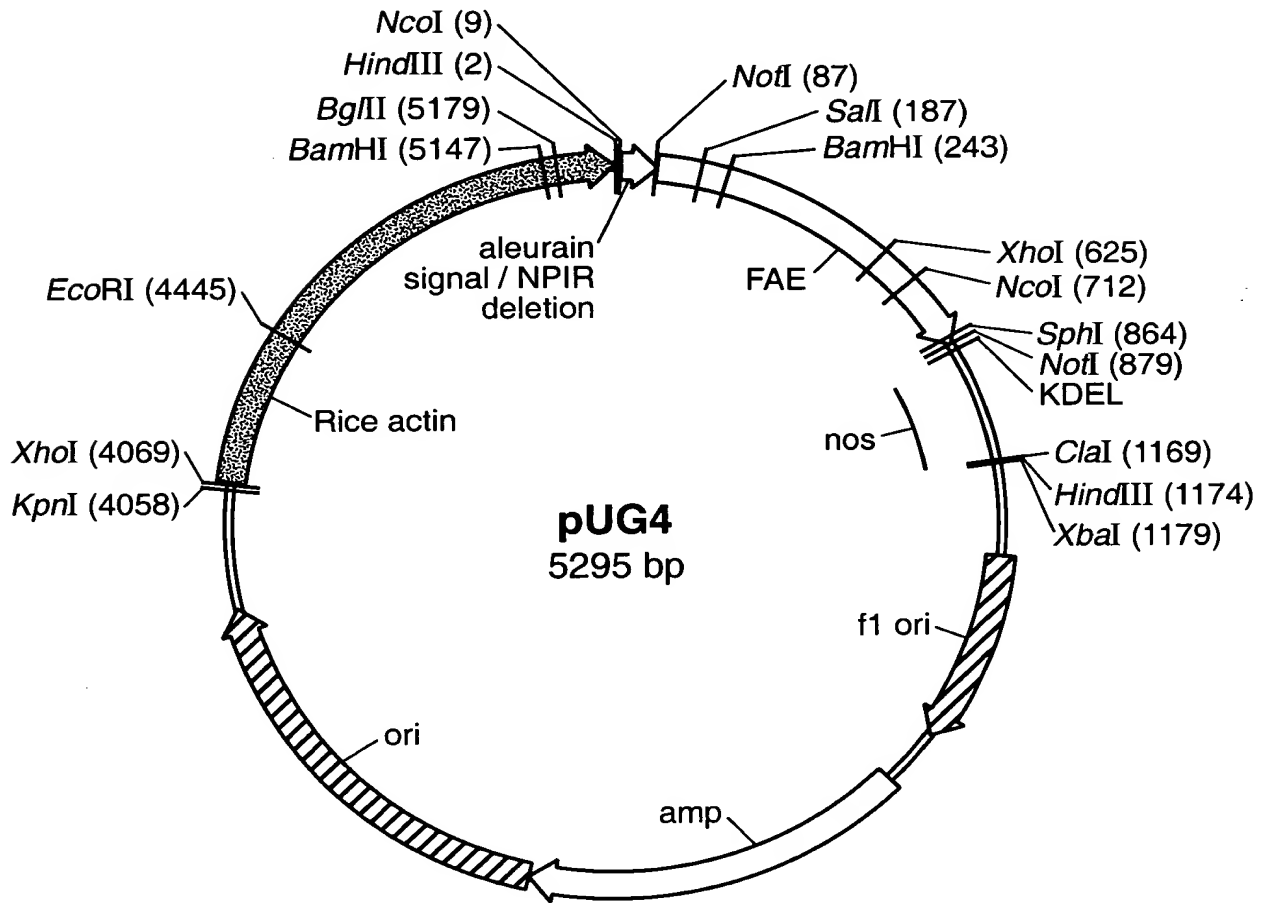
4831 AGGCTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG
    TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC TTAACACTCG CCTATTGTTA AAGTGTGTCC

                                BssHII                               EcoRI
                                ~~~~~
4901 AACAGCTAT GACCATGATT ACGCCAAGCG CGCAATTAAC CCTCACTAA GGAACAAAA GCTGG
 TTTGTGCGATA CTGGTACTAA TCGGGTTTCG CGGTTAATTG GGAGTGATTT CCCTTGTTTT CGACC

```

FIG.\_46H

132 / 154



**FIG. 47A**



133 / 154

```

NGOI
~~~~~
HindIII
~~~~~
 M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACAGC CCGGTCCTC CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCCGTCGCCG

 NotI
      ~~~~~
      . A S S R A A A S T Q G I S E D L Y S R L V E M .
71 TCGCCTCCTC CCGCGCGGCC GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT

      SalI
      ~~~~~
 . A T I S Q A A Y A D L C N I P S T I I K G E K
141 GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCGTCGA CTATTATCAA GGGAGAGAAA

 BamHI
      ~~~~~
      I Y N S Q T D I N G W I L R D D S S K E I I T V
211 ATTACAATT CTCAAACTGA CATTACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAATCACCG
      . F R G T G S D T N L Q L D T N Y T L T P F D T .
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAATCGA TACTAACTAC ACCCTCACGC CTTTCGACAC
      . L P Q C N G C E V H G G Y Y I G W V S V Q D Q
351 CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA
      V E S L V K Q Q V S Q Y P D Y A L T V T G H X L
421 GTCGAGTCGC TTGTCAAACA GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCCACKCCC
      . G A S L A A L T A A Q L S A T Y D N I R L Y T .
491 TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TCGACATAC GACAACATCC GCCTGTACAC

      XhoI
      ~~~~~
 . F G E P R S G N Q A F A S Y M N D A F Q A S S
561 CTTGGCGGAA CCGCGCAGCG GCAATCAGGC CTTGCGTCTG TACATGAACG ATGCTTCCA AGCCTCGAGC
 P D T T Q Y F R V T H A N D G I P N L P P V E Q
631 CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC CCGGTGGAGC

 NgoI
      ~~~~~
```

FIG..47B

134 / 154

```

      . G Y A H G G V E Y W S V D P Y S A Q N T F V C .
701 AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCAGAAACA CATTGTGCTG
      . T G D E V Q C C E A Q G G Q G V N A H T T Y
771 CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT

      SphI                               NotI
      ~~~~~                               ~~~~~
 F G M T S G A C T W P V A A A E P L K D E L *
841 TTTGGGATGA CGAGCGGCGC ATGCACCTGG CCGGTCGCGG CCGCGGAACC ACTGAAGGAT GAGCTGTAAA
911 GAAGCAGATC GTTCAAACAT TTGGCAATAA AGTTCTTAA GATTGAATCC TGTGCGCGGT CTTGCGGATGA
981 TTATCATATA ATTTCTGTTG AATTACGTTA AGCATGTAAT AATTAACATG TAATGCATGA CGTTATTATAT
1051 GAGATGGGTT TTTATGATTA GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAAACAA AATATAGCGC

 HindIII
      ~~~~~

1121 GCAAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT TACTAGATCG ATAAGCTTCT AGAGCGGCGC
1191 GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTACGGC CGCTCACTGG CCGTCGTTTT ACAACGTCGT
1261 GACTGGGAAA ACCCTGGCGT TACCCAACCTT AATCGCCTTG CAGCACATCC CCTTTTCGCC AGCTGGCGTA
1331 ATAGCGAAGA GCGCCGCACC GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGGACGCGCC
1401 CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAGCGCC
1471 CTAGCGCCCG CTCCTTTTCG CTCTCTCCCT TTTCTTCCCT TCCCTTTCG CACAGTTCGC CGCTTTCCC CGTCAAGCTC
1541 TAAATCGGGG GCTCCCTTTA GGGTTCGGAT TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA
1611 GGGTGATGGT TCACGTAGTG GGCCATCGCC CTGATAGACG GTTTTTTCGCC CTTTGACGTT GGAGTCCACG
1681 TTCTTTAATA GTGGACTCTT GTTCCAAACT GGAACAACAC TCAACCCCTAT CTCGGTCTAT TCTTTTGATT
1751 TATAAGGGAT TTTGCCGATT TCGGCCCTATT GGTAAAAAA TGAGCTGATT TAACAAAAAT TTAACGCGAA
1821 TTTTAAACAA ATATTACGC TTACAATTTA GGTGGCACTT TTCGGGGAAA TGTGCGCGGA ACCCCTATTT
1891 GTTTATTTTT CTAAATACAT TCAATATATG ATCCGCTCAT GAGACAATAA CCCTGATAAA TGCTTCAATA
1961 ATATTGAAA AGGAAGAGTA TGAGTATTCA ACATTTCCTG GTCCGCCCTTA TTCCCTTTTT TGCGGCATTT
2031 TGCCTTCCTG TTTTGTCTCA CCCAGAAAAG CTGGTGAAG TAAAGATGC CTTGAGAGT TTTGCGCCCG AAGAACGTTT
2101 GAGTGGGTTA CATCGAATG GATCTCAACA GCGGTAAGAT CTTGAGAGT CTTGAGAGT TTTGCGCCCG AAGAACGTTT
2171 TCCAATGATG AGCACTTTTA AAGTTCTGCT ATGTGCGCGG GTATTATCCC GTATTGACGC CGGGCAAGAG
2241 CAACTCGGTC GCCGATACA CTATTCTCAG AATGACTTGG TTGAGTACTC CATAACCATG AGTGATAACA CTGCGGCCAA
2311 TTACGGATGG CATGACAGTA AGAGAATTAT GCAGTGCTGC CATAACCATG AGTGATAACA CTGCGGCCAA
2381 CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAACC GCTTTTTCG ACAACATGGG GGATCATGTA
2451 ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAAAACGA CGAGCGTGAC ACCACGATGC
2521 CTGTAGCAAT GGCAACAACG TTGCGCAAC TATTAACTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA

```

FIG.-47C

135 / 154

```

2591 ATTAATAGAC TGGATGGAG CGGATAAAGT TGCAGGACCA CTTCCTGCGCT CGCCCTTCC GCGTGGCTGG
2661 TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG
2731 GTAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA
2801 GATCGCTGAG ATAGGTGCTT TAAAACCTCA TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGTATCTT
2871 TAGATTGATT TAAAACCTCA TTTTCGTTCC ACTGAGCGTC AGACCCCGTA AAGTATCTT AATCTCATGA
2941 CCAAAATCCC TTAACGTGAG TTTTCGTTCC ACTGAGCGTC AGACCCCGTA AAGTATCTT AAGTATCTT
3011 TTGAGATCCT TTTTCGTTCC ACTGAGCGTC AGACCCCGTA AAGTATCTT AAGTATCTT AAGTATCTT
3081 TGTTTGCCCG ATCAAGAGCT ACCAATCTTT CCGTAGTTAG GCGGCTGAA TACCTGGCTT CAAGAACTCT GTAGCACTG
3151 ATACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT
3221 CGCTCTGCTA ATCTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT TACTGTCTT
3291 AGACGATAGT TACCGGATAA GCGGCTGAA TACAGCGTGA CAGGCTCGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG
3361 AGCGAACGAC CTACACCGAA CTGAGATACC GCGGCTGAA TACAGCGTGA CAGGCTCGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG
3431 GAGAAAGGCG GACAGGTATC CCGTAAAGCG TAGTCCCTGTC GGGTTCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG
3501 GGAACGCTT GGTATCTTTA TAGTCCCTGTC GGGTTCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG
3571 GCTCGTCAGG GGGCGGAGC CTATGGAATA CCGGCTGTC GGGTTCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG
3641 CTGGCTTTT GCTCACATGT TCTTCTCTGC CCGCAGCCGA ACAGCGGAGC GCGGCTGTC GGGTTCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG
3711 GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACAGCGGAGC GCGGCTGTC GGGTTCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG
3781 AGCGCCCAAT ACGCAACCG CCGTCTCCCG CCGGCTGTC GGGTTCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG
3851 TCCCGACTGG AAAGCGGCA GTGAGCGCA CCGAATTAAT GTGAGTTAGC TCACTCATTA GGCACCCAG
3921 GCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGA TTTGTGAGCGG ATAACAATTT CACACAGGAA

3991 ACAGCTATGA CCATGATTAC GCCAAGCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG

4061 GCGCCCTCTC GAGGTCATTC ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAAC AAAGGTAAGA
4131 TTACCTGGTC AAAAGTGAAA ACATCAGTTA AAAGGTGGTA TAAGTAAAT ATCGGTAATA AAAGGTGGCC
4201 CAAAGTGAAA TTTACTCTTT TCTACTATTA TAAAATTTGA GGATGTTTGG TCGGTACTTT GATACGTCAT
4271 TTTTGTATGA ATTGGTTTTT AAGTTTATTC GCGATTGGA AATGCATATC TGTTATTTAG TCGGTTTTTA
4341 AGTTCTGTTG TTTTGTAAAT ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAAACCCAT ATGCTAATTT

4411 GACATAAATT TTGAGAAAAA TATATATTCA GCGGAATTCC ACAATGAACA ATAATAAGAT TAAAAATAGCT
4481 TGCCCCCGTT GCAGCGATGG GTATTTTCTC TAGTAAATAA AAAGATAAAC TTAGACTCAA AACATTTACA
4551 AAAACAACCC CTAAAGTCTT AAAGCCCAAA GTGCTATGCA CGATCCATAG CAAGCCAGC CCAACCCCAAC
4621 CCAACCCCAAC CCACCCCACT GCAGCCCACT GGCAATAGT CTCCACCCCT GGCATATCA CCGTGAGTTG

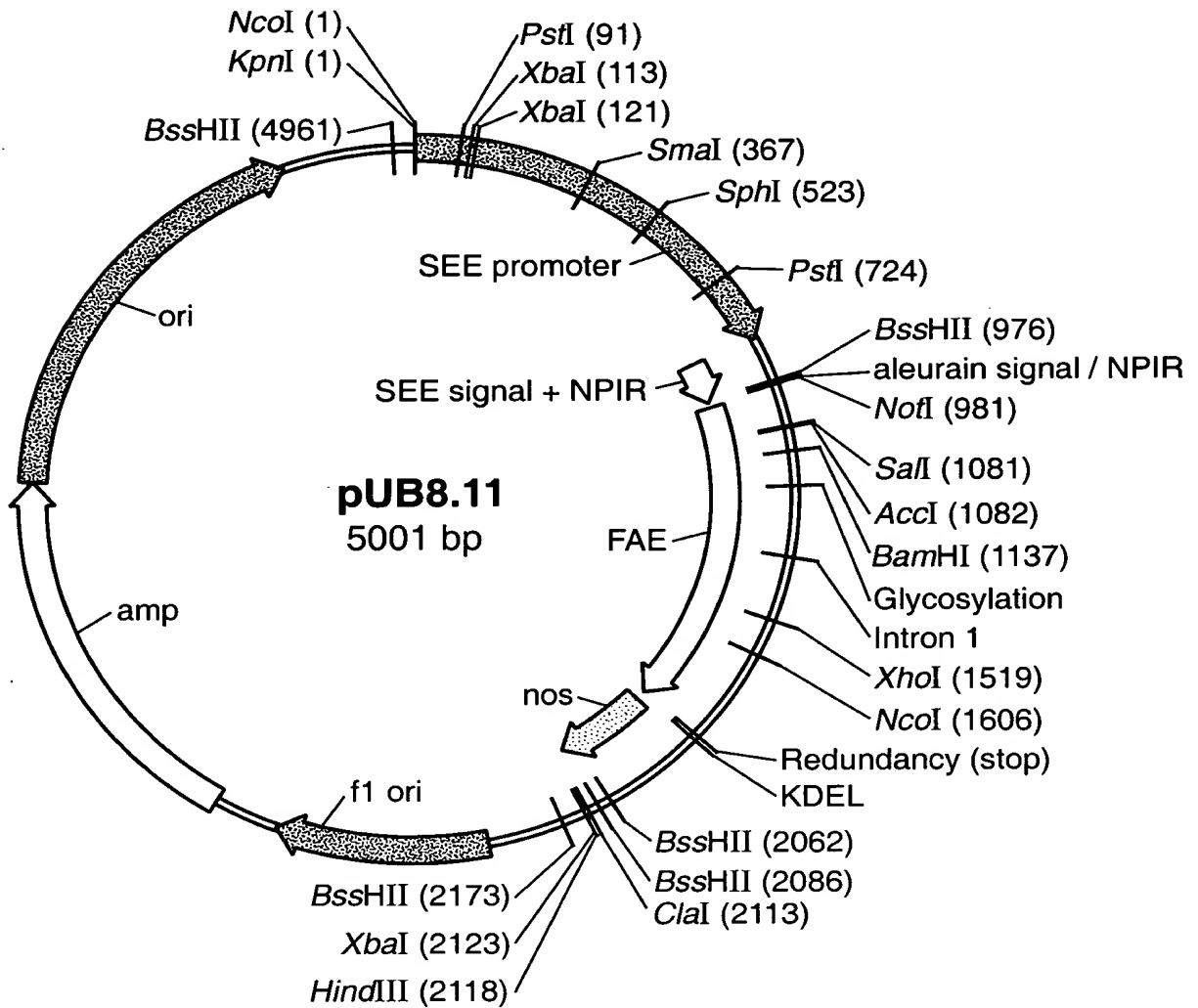
```

FIG.\_47D

4691 TCCGCACCCAC CGCACGTCTC GCAGCCAAAA AAAAAGG AAGAAAAAGA AAAACAGCAG  
4761 GTGGGTCCGG GTCGTGGGG CCGGAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC  
4831 CGCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCCAACCC  
4901 TACCACCCACC ACCACCACCA CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCTCCCCC CTCCTCCCTCC  
4971 GCCGCCGCCG GTAACCAACC CGCCCCCTCT CTCCTTCTTT CTCCGTTTTT TTTTTCGTCT CCGTCTCGAT  
5041 CTTTGGCCCTT GGTAGTTTGG GTGGGCGAGA GCGGCTTCGT CGCCACAGATC GGTGCGCGGG AGGGCGCGGA  
BamHI BglII  
~~~~~  
5111 TCTCGCGGCT GCGGTCCTCC GCGGTGAGTC GGCCCGGATC CTCGCGGGGA ATGGGGCTCT CGGATGTAGA
BglII
~~~  
5181 TCTTCTTTCT TTCTTCTTTT TGTGGTAGAA TTTGAATCCC TCAGCATGTG TCATCGGTAG TTTTTCCTTT  
5251 CATGATTGT GACAAAATGCA GCCTCGTGCG GAGCTTTTT GTAGC

FIG.\_47E

137 / 154



**FIG.\_48A**

138 / 154

```

NcoI
~~~~~
KpnI
~
1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
 GTACCCGGTC CATATTAATA CCTATAGAG TTCGTTTATT AGCTTTATAG TGGTAACCGA TGTTATAGAC

 PstI
                                     ~~~~~
71  AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
   TCGAGGCTCA AGACTGACGT CAGACCTACT CAGCACAACA TAGATCTTGA GATCTATCGT GTCGGTGTCTG

                                     XbaI
                                     ~~~~~
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTTAC CTCCTGACGT
 TGGATGTCCT CACGCTGTGA ACACCTGACA TCATCACAAC CTCTGCCTCG AGAAAGGATG GAGGACTGCA

211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCCTCAACCAA TCACGCGCTC CCAACAATAAT ATCGTCCCCC
 ACGCGGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGC GCGAG GGTGTGTTTA TAGCAGGGGG

281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGCGC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACACA CTTAGAGCGA AGGTGACCGG

 SmaI
                                     ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCTG TCACCCCTGG CGTCATGGGA
   TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCCT

421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
   ACCTTTTCTT GGAGGCAACG AGCCTACTCA GTCGGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA

                                     SphI
                                     ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATTGAC TTTATTTTGG
 AGTCTAAGGG GGTATATCTCT TTCATATCGT ACGAAAGCCC AAAACAACCC GAATTAACTG AAATAAAAAAC

561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
 AACCTCAACT TACGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCCTA TTATCCGACC

```

FIG.\_48B

139 / 154

```
631 CTAATTAATT TATAGCAAGA TTCGTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

PstI
~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAAGTG TGAAGCGTGG CGATGAGATG GGTATAAAAC
AGTAGTCGGA CTTGTGAGA CGTCTCGGAC TTCGTTCAAC ACTTCGCACC GCTACTCTAC CCATATTTTG

771 CCCCCGCACC GGGACGCGAG CTCGCCCTA CCAGTACCAT CTCGCCCTCG TCCCCCTGCC GGACGACCCA
GGGGCCGTGG CCTGCGCTC GAGGCGGAT GGTCAATGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGCCCAAG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
CATTTTATGA CAACGGGTGA GCGGCCGCTC TACCGGGTGC CGCGGTAGGA GAAGAACC GC GAGCGGCAGA

BssHII
~~~~~
NotI
~~
911 TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
ACCGGTGGCG GCGCCACCGG CGCGGTAGNA ACCGCCGTAG GTTGGCTAG GCCGGGCAGT GGCTCGCGCG

NotI
~~~~~
981 GGCCGCCCTCC ACGCAGGGCA TCTCCGAAGA CCTCTACAGC CGTTTAGTGC AAATGGCCAC TATCTCCCAA
CCGGCGGAGG TCGGTCCCGT AGAGGCTTCT AGAGGCTGCG GCAAATCAGC TTACCCGGTG ATAGAGGGTT

Sali
~~~~~
AccI
~~~~~
1051 GCTGCCCTACG CCGACCTGTG CAACATTCGG TCGACTATTA TCAAGGGAGA GAAAATTAC AATTCTCAA
CGACGGATGC GGCTGGACAC GTTGTAAGGC AGCTGATAAT AGTCCCTCT CTTTTAAATG TTAAGAGTTT
```

FIG.\_48C

140 / 154

|      |                                                                                                                                                                |  |  |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|
|      | BamHI                                                                                                                                                          |  |  |
|      | ~~~~~                                                                                                                                                          |  |  |
| 1121 | CTGACATTAA CGGATGGATC CTCCGGGACG ACAGCAGCAA AGAAATAATC ACCGTCTTCC GTGGCAGTGG<br>GACTGTAATT GCCTACCTAG GAGGCGCTGC TGTCTGTCGTT TCCTTATTAG TGGCAGAAGG CACCGTGACC  |  |  |
| 1191 | TAGTGATACG AATCTACAAC TCGATACTAA CTACACCCCTC AGCCTTTTCG ACACCCCTACC ACAATGCAAC<br>ATCACTATGC TTAGATGTTG AGCTATGATT GATGTGGGAG TCGGGAAGC TGTGGGATGG TGTACGTTG   |  |  |
| 1261 | GGTTGTGAAG TACACGGTGG ATATTATATT GGATGGGTCT CCGTCCAGGA CCAAGTCGAG TCGCTTGTCA<br>CCAACACTTC ATGTGCCACC TATAATATAA CCTACCCAGA GGCAGGTCCT GGTTCAGCTC AGCGAACAGT   |  |  |
| 1331 | AACAGCAGGT TAGCCAGTAT CCGGACTACG CGCTGACCCGT GACCGGCCAC KCCCTCGGCG CCTCCCTGGC<br>TTGTCTGTTCA ATCGGTCATA GGCTTGATGC GCGACTGGCA CTGGCCGGTG MGGGAGCCGC GGAGGGACCG |  |  |
| 1401 | GGCACTCACT GCCGCCCAGC TGCTTGGGAC ATACGACAAC ATCCGCCCTGT ACACCTTCGG CGAACCCGCGC<br>CCGTGAGTGA CGGCGGGTCG ACAGACGCTG TATGCTGTTG TAGGCGGACA TGTGGAAGCC GCTTGGCGCG |  |  |
|      | XhoI                                                                                                                                                           |  |  |
|      | ~~~~~                                                                                                                                                          |  |  |
| 1471 | AGCGGCAATC AGGCCTTCGC GTCGTACATG AACGATGCCT TCCAAGCCTC GAGCCCAGAT ACGACGCAGT<br>TCGCCCCTAG TCCGGAAGCG CAGCATGTAC TTGCTACGGA AGGTTCCGGAG CTCGGGTCTA TGCTGCCGTC  |  |  |
|      | NcoI                                                                                                                                                           |  |  |
|      | ~~~~~                                                                                                                                                          |  |  |
| 1541 | ATTTCGGGT CACTCATGCC AACGACGGCA TCCCAAACCT GCCCCCGGTG GAGCAGGGGT ACGCCCATGG<br>TAAAGGCCCA GTGAGTACGG TTGCTGCCGT AGGGTTTGA CGGGGGCCAC CTCGTCCCCA TCGGGGTACC     |  |  |
| 1611 | CGGTGTAGAG TACTGGAGCG TTGATCCTTA CAGCGCCAG AACACATTG TCTGCACCTGG GGATGAAGTG<br>GCCACATCTC ATGACCTCGC AACTAGGAAT GTCGCGGGTC TTGTGTAAAC AGACGTGACC CCTACTTCAC    |  |  |
| 1681 | CAGTGCTGTG AGGCCCAGGG CGGACAGGGT GTGAATAATG CGCACACGAC TTATTTTGGG ATGACGAGCG<br>GTCACGACAC TCCGGGTCCC GCCTGTCCCA CACTTATTAC GCGTGTCTG AATAAAACCC TACTGCTCGC    |  |  |
| 1751 | GAGCCTGTAC ATGGTGATCA GTCATTTTCAG CCTCCCCGAG TGTACCAGGA AAGATGGATG TCCTGGAGAG<br>CTCGGACATG TACCACCTAGT CAGTAAAGTC GGAGGGGCTC ACATGGTCTT TCTACCTAC AGGACCTCTC  |  |  |

FIG.\_48D



141 / 154

```

1821 GGGGCCGCGT AACCACTGAA GGATGAGCTG TAAAGAAGCA GATCGTTCAA ACATTGGGCA ATAAAGTTTC
      CCCC GGCGCA TTGGTGACTT CCTACTCGAC ATTTCTTCGT CTAGCAAGTT TGTAAACCGT TATTCAAAG

1891 TTAAGATTGA ATCCTGTTGC CGGTCTTGGC ATGATTATCA TATAATTTCT GTTGAATTAC GTTAAGCATG
      AATCTAACT TAGGACAAAC GCCAGAACGC TACTAATAGT ATATTAAAGA CAACTTAATG CAATTCGTAC

1961 TAAATAATTAA CATGTAATGC ATGACGTTAT TTATGAGATG GGTTTTATG ATTAGAGTCC CGCAATTATA
      ATTATTAATT GTACATTACG TACTGCAATA AATACTCTAC CAAAATAAC TAATCTCAGG GCGTTAATAT

      BssHII          BssHII
      ~~~~~          ~~~~~

2031 CATTTAATAC GCGATAGAAA ACAAATATA GCGCGCAAAC TAGGATAAAT TATCGCGCGC GGTGTCATCT
 GTAAATTATG CGCTATCTTT TGTTTTATAT CGCGCGTTTG ATCCTATTTA ATAGCGCGCG CCACAGTAGA

 XbaI
      ~~~~~

      ClaI HindIII
      ~~~~~

2101 ATGTTACTAG ATCGATAAGC TTCTAGAGCG GCCGGTGGAG CTCCAATTCG CCCTATAGTG AGTCGTATTA
 TACAATGATC TAGCTATTTC AAGATCTCGC CGGCCACCTC GAGGTTAAGC GGGATATCAC TCAGCATAAT

 BssHII
      ~~~~~

2171 CGCGCGCTCA CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAACCCCTG GCGTTACCCA ACTTAATCGC
      GCGCGCGAGT GACCGGCAGC AAAATGTTGC AGCACTGACC CTTTGGGAC CGCAATGGGT TGAATTAGCG

2241 CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC
      GAACGTCGTG TAGGGGGAAA CGGTCGACC GCATTATCGC TTCTCCGGGC GTGGCTAGCG GGAAGGGTTG

2311 AGTTGCGCAG CCTGAATGGC GAATGGGACG CGCCCTGTAG CGGCGCATT AAGCGCGCGG GTGTGGTGGT
      TCAACGCGTC GGACTTACCG CTTACCCTGC GCGGGACATC GCCGCGTAAT TCGCGCCGCC CACACCAACA

2381 TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTCTTT CCCTTCCTTT
      ATGCGCGTCG CACTGGCGAT GTGAACGGTC GCGGGATCGC GGGCGAGGAA AGCGAAAGAA GGAAGGAAA

```

FIG. 48E

142 / 154

2451 CTCGCCACGT TCGCCGGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG  
GAGCGGTGCA AGCGGCCGAA AGGGCAGTT CGAGATTTAG CCCCCGAGG AATCCCAAG GCTAAATCAC

2521 CTTTACGGCA CCTCGACCCC AAAAAACTTG ATTAGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA  
GAAATGCCGT GGAGCTGGGG TTTTGTGAAC TAATCCCACT ACCAAGTGCA TCACCCGGTA GCGGGACTAT

2591 GACGGTTTT CGCCCTTTGA CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTCCA AACTGGAACA  
CTGCCAAAA GCGGAACT GCAACCTCAG GTGCAAGAAA TTATCACCTG AGAACCAAGGT TTGACCTTGT

2661 ACACCTCAACC CTATCTCGGT CTATCTTTT GATTTATAAG GGATTTTGCC GATTCGGCC TATTGGTTAA  
TGTGAGTTGG GATAGAGCCA GATAAGAAA CTAAATATTC CCTAAAAACGG CTAAAGCCGG ATAACCAATT

2731 AAAATGAGCT GATTTAACAA AAATTTAAG CGAATTTTAA CAAAATATTA ACGCTTACAA TTTAGGTGGC  
TTTTTACTCGA CTAAATTGTT TTTAAATTGC GCTTAAAAAT GTTTTATAAT TCGGAATGTT AAATCCACCG

2801 ACTTTTCGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT ATGTATCCCG  
TGAAAAAGCCC CTTTACACGC GCCTTGGGA TAAACAAATA AAAAGATTTA TGTAAGTTTA TACATAGGCG

2871 TCATGAGACA ATAACCCCTGA TAAATGCTTC AATAATATG AAAAAGGAAG AGTATGAGTA TTCAACATTT  
AGTACTCTGT TATTGGGACT ATTTACGAAG TTATTATAAC TTTTTCCTTC TCATACTCAT AAGTTGTAAA

2941 CCGTGTCCGC CTTATTCCCT TTTTTCGGC ATTTTGCCCTT CCTGTTTTG CTCACCCAGA AACGTGGTG  
GGCACAGCG GAATAAGGA AAAAACGCCG TAAACGGAA GGACAAAAAC GAGTGGGTCT TTGCGACCAC

3011 AAAGTAAAG ATGCTGAAGA TCAGTTGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA  
TTTCATTTTC TACGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT TGACCTAGAG TTGTCCGCCAT

3081 AGATCCTTGA GAGTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTTAAAGTTC TGCTATGTGG  
TCTAGGAAT CTCAAAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA AAATTTCAAG ACGATACACC

3151 CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTCCGCCGA TACACTATTC TCAGAATGAC  
GCGCCATAAT AGGCATAAC TGCGGCCCGT TCCTCGTTGAG CCAGCGCGCT ATGTGATAAG AGTCTTACTG

3221 TTGGTTGAGT ACTCACCAGT CACAGAAAA GATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG  
AACCAACTCA TGAGTGGTCA GTGTCTTTTC GTAGAATGCC TACCGTACTG TCATTCTCTT AATACGTAC

FIG.\_48F

143 / 154

```

3291 CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAAACG ATCGGAGGAC CGAAGGAGCT
    GACGGTATTG GTACTCACTA TTGTGACGCC GGTGAATGA AGACTGTTGC TAGCCTCCCTG GCTTCCTCGA

3361 AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACCTCGC CTTGATCGTT GGAACCCGGA GCTGAATGAA
    TTGGCGAAAA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAAC TAGCAA CCCTTGGCCT CGACTTACTT

3431 GCCATACCAA ACGACGAGCG TGACACCACG ATGCCCTGTAG CAATGGCAAC AACGTTGCGC AAAC TATTAA
    CCGTATGGTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACCCTTG TTGCAACGCG TTTGATATAAT

3501 CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAAATTAAT AGACTGGATG GAGGCGGATA AAGTTGCAGG
    GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCCGCCCTAT TTCAACGTCC

3571 ACCACTTCTG CGCTCGGCCC TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG
    TGGTGAAAGAC GCGAGCCGGG AAGGCCGACC GACCAATAA CGACTATTA GACCTCGGCC ACTCGCACCC

3641 TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG
    AGAGCGCCAT AGTAACGTCG TGACCCCGGT CTACCATTCTG GGAGGGCATA GCATCAATAG ATGTGCTGCC

3711 GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG
    CCTCAGTCCG TTGATACCTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA CGGAGTGACT AATTCGTAAC

3781 GTAACGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTAAAAAC TTCA TTTTAA AAGG
    CATTGACAGT CTGGTTCAAA TGAGTATATA TGAAATCTAA CTAAATTTTG AAGTAAAAAT TAAATTTTCC

3851 ATCTAGGTGA AGATCCTTTT TGATAAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG TTCCACTGAG
    TAGATCCACT TCTAGGAAAA ACTATTAGAG TACTGGTTT AGGGAATTGC ACTCAAAAGC AAGGTGACTC

3921 CGTCAGACCC CGTAGAAAAA ATCAAAAGGAT CTTCTTTGAGA TCCTTTT TTTT CTGCGCGTAA TCTGCTGCTT
    GCAGTCTGGG GCATCTTTTC TAGTTTCCCTA GAAGAACTCT AGGAAAAAAA GACGCGCATT AGACGACGAA

3991 GCAAAACAAA AAACCAACCGC TACCAGCGGT GGTTTGTGTA CCGGATCAAG AGCTACCAAC TCTTTTTCGG
    CGTTTGT TTTGGTGGCG ATGGTCGCCA CCAACCAAAAC GGCTTAGTTC TCGATGGTTG AGAAAAAGGC

4061 AAGGTAAC TGCTTCAGCAG AGCGCAGATA CCAAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC
    TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTTTATGAC AGGAAGATCA CATCGGCATC AATCCGGTGG
  
```

FIG. 48G

144 / 154

4131 ACTTCAAGAA CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG  
TGAAGTTCTT GAGACATCGT GCGGATGTA TGGAGCGAGA CGATTAGGAC AATGGTCACC GACGACGGTC

4201 TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG ATAAGGCGCA GCGGTCGGGC  
ACCGCTATT CAGCAGAAAT GGGCCAACTT GAGTTCTGCT ATCAATGGCC TATTCCGCGT CGCCAGCCCG

4271 TGAACGGGG GTTCGTGCAC ACAGCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC  
ACTTGCCCC CAAGCACGTG TGTCGGGTG AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGATGTG

4341 GTGAGCTATG AGAAAGCGC ACGCTTCCC AGGGGAAAC GCCTGGTATC TTATATAGTCC TGTCGGGTTT  
CACTCGATAC TCTTTCGGG TCGAAGGC TTCCCTCTTT CGGACCATAG AAATATCAGG ACAGCCCAA

4411 CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGAAAC GCCTGGTATC TTATATAGTCC TGTCGGGTTT  
GCCTTGTCCT CTCGCGTGCT CCTCGAAGG TCCCCCTTTG CGGACCATAG AAATATCAGG ACAGCCCAA

4481 CGCCACCTCT GACTTGAGCG TCGATTTTG TGATGCTCGT CAGGGGGCG GAGCCTATGG AAAAAAGCA  
GCGGTGGAGA CTGAACTCGC AGCTAAAAAC ACTACGAGCA GTCCCCCGC CTCGGATACC TTTTTCGGGT

4551 GCAACGCGG CTTTTTACGG TTCCCTGGCT TTTGCTGCC TTTTGCTCAC ATGTTCTTTC CTGCGTTATC  
CGTTGCGCG GAAAAATGCC AAGGACCGA AAACGACCGG AAAACGAGTG TACAAGAAAG GACGCAATAG

4621 CCCTGATTCT GTGGATAACC GTATTACCG CTTTGAGTGA GCTGATACCG CTCGCCGCGC CCGAACGACC  
GGGACTAAGA CACCTATTGG CATAAATGGG GAAACTCACT CGACTATGGC GAGCGGCGTC GGTTCGTGG

4691 GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGC CAATACGCAA ACCGCCCTCT CCCGCGCGTT  
CTCGCGTCGC TCAGTCACTC GTCCTTCGC CTTCTCGCG GTTATGCGTT TGGCGGAGAG GGGCGCGCAA

4761 GGCCGATTCA TTAATGCAGC TGGCACGACA GGTTCCTCGA CTGGAAGCG GGCAGTGAGC GCAACGCAAT  
CCGGCTAAGT AATTACGTG ACCGTGCTGT CCAAAGGGCT GACCTTTCCG CCGTCACCTCG CGTTGCGTTA

4831 TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC TTCCGGCTCG TATGTTGTGT  
ATTACACTCA ATCGAGTGAG TAATCCGTGG GGTCCGAAAT GTGAAATACG AAGGCCGAGC ATACAACACA

FIG.\_48H

4901 GGAATTGTGA GCGGATAACA ATTTACACACA GGAACACAGCT ATGACCATGA TTACGCCAAG CGCGCAATTA  
CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTGTGCGA TACTGGTACT AATGCGGTTC GCGCGTTAAT

BssHII  
~~~~~

4971 ACCCTCACTA AAGGGAACAA AAGCTGGGTA C
TGGGAGTGAT TTCCCTTGTT TTCGACCCAT G

NcoI
~~~~~

KpnI  
~~~~~

FIG._48I

146 / 154

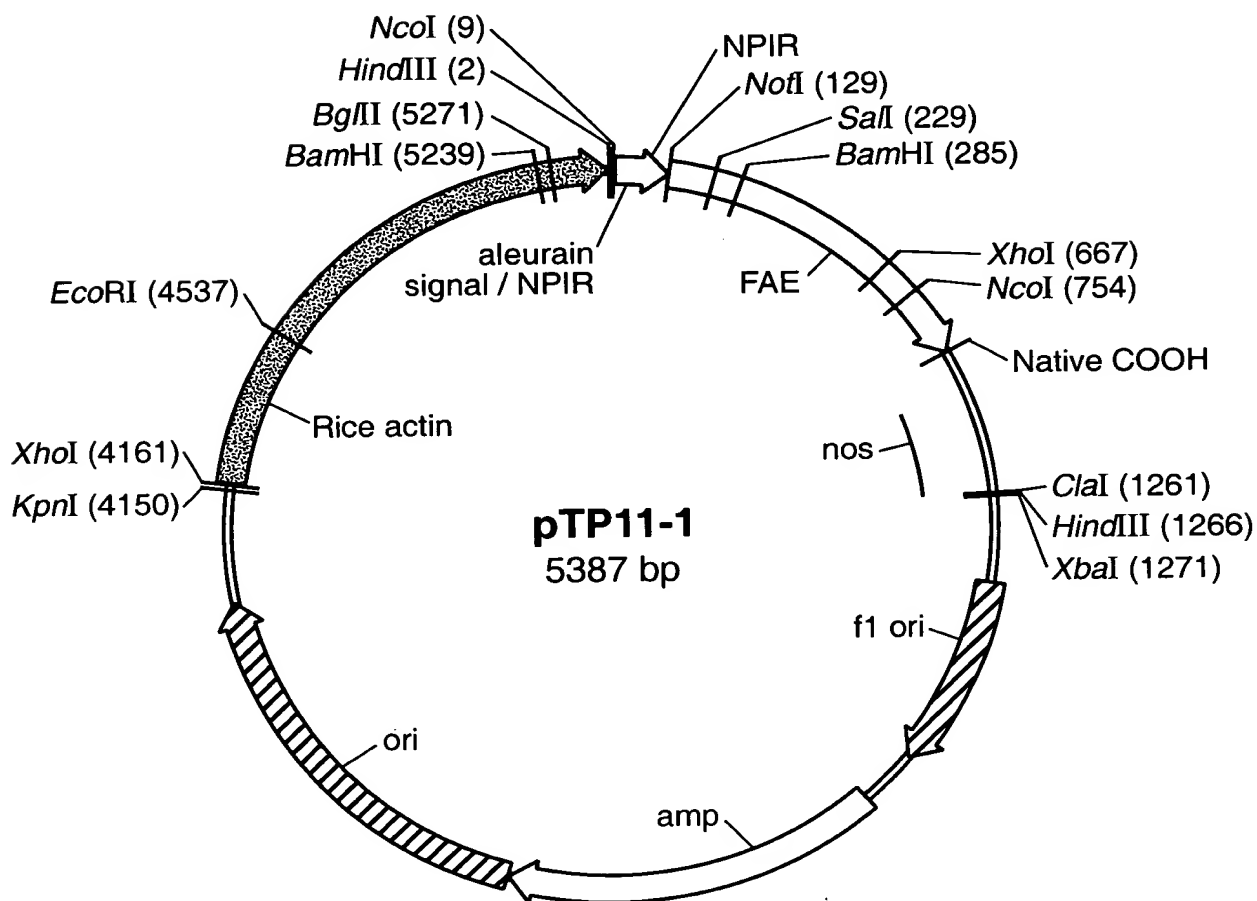


FIG._49A

147 / 154

```

NcoI
~~~~~
HindIII
~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1  AAGCTTACCA TGGCCACAGC CCGCGTCCTC CTCCTGGGCG TCGCCGTGCT GGCACGGCC GCCGTCGCCG
                                NotI
                                ~~~~~
      . A S S S F A D S N P I R P V T D R A A A S T .
71  TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCCGATCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC
      . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
                                SalI
                                ~~~~~
                                ACCI
                                ~~~~~
      D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTTACAA TTCTCAAACT GACATTAACG
                                BamHI
                                ~~~~~
      . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACCTGGTA GTGATACGAA
      . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAACTC GATACTAACT ACACCCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
      H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA
      . Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
      . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGGCGAG CGGCAATCAG
                                XhoI
                                ~~~~~
      A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA
```

FIG. 49B

148 / 154

```

      . H A N D G I P N L P P V E Q G Y A H G G V E Y .
701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA
    . W S V D P Y S A Q N T F V C T G D E V Q C C E
771 CTGGAGCGTT GATCCTTACA GCGCCACAGAA CACATTGTGC TGCACCTGGG ATGAAGTGCA GTGCTGTGAG
    A Q G G Q G V N N A H T T Y F G M T S G A C T W
841 GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGA GCCTGTACAT
    *
911 GGTGATCAGT CATTTAGCC TCCCCGAGTG TACCAGGAAA GATGGATGTC CTGGAGAGGG GGCCGCGTAA
981 CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT
1051 CCTGTGCGG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT TAAGCATGTA ATAATTAAACA
1121 TGTAAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG CAATTATACA TTTAATACGC
      .
1191 GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG TGTCTATCTAT GTTACTAGAT
      HindIII
      ~~~~~
      ClaI      XbaI
      ~~~~~
1261 CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC CTATAGTGAG TCGTATTACG CGCGCTCACT
1331 GGCCGTGCTT TTACAACGTC GTGACTGGGA AACCCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT
1401 CCCCCTTTTC CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC
1471 TGAATGGCGA ATGGGACGCG CCTGTAGCG CGCTCCTTTC GCTTCTTCC GTGGTGGTTA CGCGCAGCGT
1541 GACCGCTACA CTTGCCAGCG CCTAGCGCC CGCTCCTTTC GCTTCTTCC GTTCCCTTCT CGCCACGTTT
1611 GCCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC
1681 TCGACCCCAA AAACTTGAT TAGGTGATG TAGGTGATG TGGGCCATCG CCCTGATAGA CCGTTTTCG
1751 CCCTTTGACG TTGGAGTCCA CGTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAACT ACTCAACCCCT
1821 ATCTCGGTCT ATCTTTTGA TTTATAAGGG ATTTTGGCGA TTTCCGCCCTA TTGGTTAAAA AATGAGCTGA
1891 TTTAACAAAA ATTTAACGCG AATTTTAACTA AATATTAAC GCTTACAATT TTAGTGGCAC TTTTCGGGGA
1961 AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT TATCCGCTC ATGAGACAAT
2031 AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTCGCCCT
2101 TATTCCTTTT TTTGCGGCAT TTTGCCCTTC TGTTTGTGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT
2171 GCTGAAGATC AGTTGGGTGC ACAGTGGGT TTTCCAATGA TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA
2241 GTTTTCGCCC CGAAGAACGT TTTCCAATGA TGAGCACTTT TAAAGTTCTG TAAAGTTCTG CCGTATTATC
2311 CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC
2381 TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA

```

FIG. 49C

2451	TGAGTGATAA	CACTGCGGCC	AACCTACTTC	TGACAAACGAT	CGGAGGACCG	AAGGAGCTAA	CCGCTTTTTT
2521	GCACAACATG	GGGATCATG	TAACTCGCCT	TGATCGTTGG	GAAACCGGAGC	TGAATGAAGC	CATACCAAAC
2591	GACGAGCGTG	ACACCACGAT	GCCTGTAGCA	ATGGCAACAA	CGTTGCGCAA	ACTATTAACT	GGCGAACTAC
2661	TTACTCTAGC	TTCCCCGGCA	CAATTAATAG	ACTGGATGGA	GGCGGATAAA	GTTGCAGGAC	CACCTCTGCG
2731	CTCGGCCCTT	CCGGCTGGCT	GGTTTATTGC	TGATAAATCT	GGAGCCGGTG	AGCGTGGGTC	TCGCGGTATC
2801	ATTGCAGCAC	TGGGGCCAGA	TGGTAAGCCC	TCCCCTATCG	TAGTTATCTA	CACGACGGGG	AGTCAGGCAA
2871	CTATGGATGA	ACGAAATAGA	CAGATCGCTG	AGATAGGTGC	CTCACTGATT	AAGCATTTGGT	AACGTGTCAGA
2941	CCAAGTTTAC	TCATATATAC	TTTAGATTGA	TTTAAAACTT	CATTTTAAAT	TTAAAAAGGAT	CTAGGTGAAG
3011	ATCCTTTTTC	ATAATCTCAT	GACCAAAATC	CCTTAAACGTG	AGTTTTCGTT	CCACTGAGCG	TCAGACCCCG
3081	TAGAAAAGAT	CAAAGGATCT	TCTTGAGATC	CTTTTTCCT	GCGCGTAATC	TGCTGCTTGC	AAACAAAAAA
3151	ACCACCGCTA	CCAGCGGTGG	TTTGTTTGCC	GGATCAAGAG	CTACCAACTC	TTTTTCCGAA	GGTAACCTGGC
3221	TTCAGCAGAG	CGCAGATACC	AAATACTGTC	CTTCTAGTGT	AGCCGTAGTT	AGGCCACCCAC	TTCAAGAACT
3291	CTGTAGCACC	GCCTACATAC	CTCGCTCTGC	TAATCCTGTT	ACCAGTGGCT	GCTGCCAGTG	CGGATAAGTC
3361	GTGCTTTACC	GGGTTGGACT	CAAGACGATA	GTTACCGGAT	AAGGCGCAGC	GGTCGGGCTG	AACGGGGGTT
3431	TCGTGCACAC	AGCCAGCTT	GGAGCGAACG	ACCTACACCG	AACTGAGATA	CCTACAGCGT	GAGCTATGAG
3501	AAAGCGCCAC	GCTTCCCGAA	GGGAGAAAGG	CGGACAGGTA	TCCGGTAAGC	GGCAGGGTCG	GAACAGGAGA
3571	GCGCACGAGG	GAGCTTCCAG	GGGAAACGC	CTGGTATCTT	TATAGTCCCTG	TCGGGTTTCG	CCACCTCTGA
3641	CTTGAGCGTC	GATTTTGTG	ATGCTCGTCA	GGGGGCGGA	GCCTATGGAA	AAACGCCAGC	AACCGGGCCT
3711	TTTACGGTT	CTTGGCCCTT	TGCTGGCCTT	TTGCTCACAT	GTTCTTTTCC	CGTTATATCC	CTGATTCTGT
3781	GGATAACCGT	ATTACCGCCT	TTGAGTGAGC	TGATACCGCT	CGCCGACGCC	GAACGACCCG	CGCAGCGGAG
3851	TCAGTGAGCG	AGGAAGCGGA	AGAGCGCCCA	ATACGCAAA	CGCCTCTCCC	CGCGGTTGG	CCGATTCAAT
3921	AATGCAGCTG	GCACGACAGG	TTTCCCGACT	GGAAAGCGGG	CAGTGAGCGC	AACGCAATTA	ATGTGAGTTA
3991	GCTCACTCAT	TAGGCACCCC	AGGCTTTACA	CTTATGCTT	CCGGCTCGTA	TGTTGTGTGG	AATTGTGAGC
4061	GGATAACAAT	TTCACACAGG	AAACAGCTAT	GACCATGATT	ACGCCAAGCG	CGCAATTAA	CCTCACTAAA
~~~~~ KpnI ~~~~~							
4131	GGGAACAAAA	GCTGGGTACC	GGGCCCCCCC	TCGAGGTCAT	TCATATGCTT	GAGAAAGAGAG	TCGGGATAGT
4201	CCAAAAATAA	ACAAAAGTAA	GATTACCTGG	TCAAAAAGTGA	AAACATCAGT	TAAAAAGGTGG	TATAAGTAAA
4271	ATATCGGTAA	TAAAAAGGTG	CCCAAGTGA	AATTACTCT	TTTCTACTAT	TATAAAAAAT	GAGGATGTTT
4341	TGTCGGTACT	TTGATACGTC	ATTTTGTAT	GAATTGGTTT	TTAAGTTTAT	TCGCGATTGG	GAAATGCATA
4411	TCTGTATTTC	AGTCGGTTTT	TAAATTTCGTT	GCTTTGTAA	ATACAGAGGG	ATTTGTATAA	GAAATATCTT
~~~~~ EcoRI ~~~~~							
4481	TAAAAAACCC	ATATGCTAAT	TTGACATAAT	TTTTGAGAAA	AATATATATT	CAGGCGAATT	CCACAATGAA
4551	CAATAATAAG	ATTAAAAATAG	CTTGCCCCCG	TTGCAGCGAT	GGGTATTTTT	TCTAGTAAAA	TAAAAAGATAA
4621	ACTTAGACTC	AAAACATTTA	CAAAAAACAAC	CCCTAAAGTC	CTAAAGCCCC	AAGTGCCTATG	CACGATCCAT

FIG. 49D

150 / 154

```
4691 AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCAACCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC
4761 CCGGCACTAT CACCGTGAGT TGTCCGCACC ACGCACGTC TCGCAGCCAA AAAAAAAA AGAAGAAAA
4831 AAAAGAAAA GAAAAACAGC AGGTGGGTCC TCGCGTGGG GGCAGGAGGA TCGCGAGCAG
4901 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATATACA TACCCCCCCC
4971 TCTCCTCCCA TCCCCCACA CCTACCACA CCACCTCCTCC CCCCTCGCTG CCCGACGACG
5041 AGCTCCTCCC CCTCCTCCC CGGCCGCCG CGGTAAACAC CCCGCCCTC TCCTCTTCT TTCTCCGTTT
5111 TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCGA GAGCGGCTTC GTCGCCCAGA

                               BamHI
                               ~~~~~

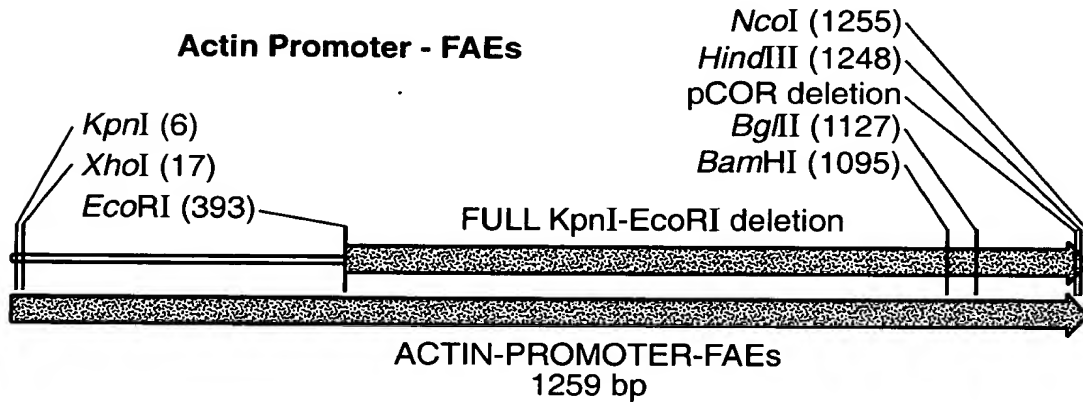
5181 TCGGTGCGCG GGAGGGGCGG GATCTCGCG GATCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG

                               BglII
                               ~~~~~

5251 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTGTGGTAG AATTGAATC CCTCAGCAT
5321 GTTCATCGGT AGTTTTCTT TTTCATGATT GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC
```

FIG._49E

151 / 154



	KpnI	XhoI					
	~~~~~	~~~~~					
1	<u>GGTACCGGGC</u>	<u>CCCCCTCGA</u>	<u>GGTCATTCAT</u>	<u>ATGCTTGAGA</u>	<u>AGAGAGTCGG</u>	<u>GATAGTCCAA</u>	<u>AATAAAACAA</u>
	<u>CCATGGCCCCG</u>	<u>GGGGGGAGCT</u>	<u>CCAGTAAGTA</u>	<u>TACGAACCTCT</u>	<u>TCTCTCAGCC</u>	<u>CTATCAGGTT</u>	<u>TTATTTTGTT</u>
71	<u>AGGTAAGATT</u>	<u>ACCTGGTCAA</u>	<u>AAGTGAAAAC</u>	<u>ATCAGTTAAA</u>	<u>AGGTGGTATA</u>	<u>AGTAAAATAT</u>	<u>CGGTAATAAA</u>
	<u>TCCATTCTAA</u>	<u>TGGACCAGTT</u>	<u>TTCACTTTTG</u>	<u>TAGTCAATTT</u>	<u>TCCACCATAT</u>	<u>TCATTTTATA</u>	<u>GCCATTATTT</u>
141	<u>AGGTGGCCCA</u>	<u>AAGTGAAATT</u>	<u>TACTCTTTTC</u>	<u>TACTATTATA</u>	<u>AAAATTGAGG</u>	<u>ATGTTTTGTC</u>	<u>GGTACTTTGA</u>
	<u>TCCACCGGGT</u>	<u>TTCACTTTAA</u>	<u>ATGAGAAAAG</u>	<u>ATGATAATAT</u>	<u>TTTAACTCC</u>	<u>TACAAAACAG</u>	<u>CCATGAAACT</u>
211	<u>TACGTCATTT</u>	<u>TTGTATGAAT</u>	<u>TGGTTTTTAA</u>	<u>GTTTATTCGC</u>	<u>GATTTGGAAA</u>	<u>TGCATATCTG</u>	<u>TATTTGAGTC</u>
	<u>ATGCAGTAAA</u>	<u>AACATACTTA</u>	<u>ACCAAAAATT</u>	<u>CAAATAAGCG</u>	<u>CTAAACCTTT</u>	<u>ACGTATAGAC</u>	<u>ATAAACTCAG</u>
281	<u>GGTTTTTAAG</u>	<u>TTCGTTGCTT</u>	<u>TTGTAAATAC</u>	<u>AGAGGGATTT</u>	<u>GTATAAGAAA</u>	<u>TATCTTTAAA</u>	<u>AAACCCATAT</u>
	<u>CCAAAAATTC</u>	<u>AAGCAACGAA</u>	<u>AACATTTATG</u>	<u>TCTCCCTAAA</u>	<u>CATATTCTTT</u>	<u>ATAGAAATTT</u>	<u>TTTGGGTATA</u>
				<u>EcoRI</u>			
				~~~~~			
351	<u>GCTAATTTGA</u>	<u>CATAATTTTT</u>	<u>GAGAAAAATA</u>	<u>TATATTCAGG</u>	<u>CGAATTCCAC</u>	<u>AATGAACAAT</u>	<u>AATAAGATTA</u>
	<u>CGATTAAACT</u>	<u>GTATTAAAAA</u>	<u>CTCTTTTTAT</u>	<u>ATATAAGTCC</u>	<u>GCTTAAGGTG</u>	<u>TTACTTGTTA</u>	<u>TTATTCTAAT</u>
421	<u>AAATAGCTTG</u>	<u>CCCCCGTTGC</u>	<u>AGCGATGGGT</u>	<u>ATTTTTTCTA</u>	<u>GTAAAATAAA</u>	<u>AGATAAACTT</u>	<u>AGACTCAAAA</u>
	<u>TTTATCGAAC</u>	<u>GGGGGCAACG</u>	<u>TCGCTACCCA</u>	<u>TAAAAAAGAT</u>	<u>CATTTTATTT</u>	<u>TCTATTTGAA</u>	<u>TCTGAGTTTT</u>
491	<u>CATTTACAAA</u>	<u>AACAACCCCT</u>	<u>AAAGTCCTAA</u>	<u>AGCCCAAAGT</u>	<u>GCTATGCACG</u>	<u>ATCCATAGCA</u>	<u>AGCCAGCCCC</u>
	<u>GTAAATGTTT</u>	<u>TTGTTGGGGA</u>	<u>TTTCAGGATT</u>	<u>TCGGGTTTCA</u>	<u>CGATACGTGC</u>	<u>TAGGTATCGT</u>	<u>TCGGGTCGGG</u>
561	<u>AACCCAACCC</u>	<u>AACCCAACCC</u>	<u>ACCCAGTGC</u>	<u>AGCCAACTGG</u>	<u>CAAATAGTCT</u>	<u>CCACCCCCGG</u>	<u>CACTATCACC</u>
	<u>TTGGGTGGG</u>	<u>TTGGGTGGG</u>	<u>TGGGGTCACG</u>	<u>TCGGTTGACC</u>	<u>GTTTATCAGA</u>	<u>GGTGGGGGCC</u>	<u>GTGATAGTGG</u>
631	<u>GTGAGTTGTC</u>	<u>CGCACCACCG</u>	<u>CACGTCTCGC</u>	<u>AGCCAAAAAA</u>	<u>AAAAAAGAA</u>	<u>AGAAAAAAA</u>	<u>GAAAAAGAAA</u>
	<u>CACTCAACAG</u>	<u>GCGTGGTGGC</u>	<u>GTGCAGAGCG</u>	<u>TCGGTTTTTT</u>	<u>TTTTTTTCTT</u>	<u>TCTTTTTTTT</u>	<u>CTTTTTCTTT</u>
701	<u>AACAGCAGGT</u>	<u>GGGTCCGGGT</u>	<u>CGTGGGGGCC</u>	<u>GGAAAAGCGA</u>	<u>GGAGGATCGC</u>	<u>GAGCAGCGAC</u>	<u>GAGGCCCGGC</u>
	<u>TTGTCGTCCA</u>	<u>CCCAGGCCCA</u>	<u>GCACCCCCCG</u>	<u>CCTTTTCGCT</u>	<u>CCTCCTAGCG</u>	<u>CTCGTCGCTG</u>	<u>CTCCGGGCCG</u>

FIG. 50A

152 / 154

771 CCTCCCTCCG CTTCCAAAGA AACGCCCCC ATCGCCACTA TATACATACC CCCCCTCTC CTCCCATCCC
GGAGGGAGGC GAAGGTTTCT TTGCGGGGG TAGCGGTGAT ATATGTATGG GGGGGGAGAG GAGGGTAGGG

841 CCCAACCCTA CCACCACCAC CACCACCACC TCCTCCCCC TCGTGCCGG ACGACGAGCT CCTCCCCCT
GGGTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA

911 CCCCCTCCGC CGCCGCCGGT AACCACCCCG CCCCTCTCCT CTTTCTTTCT CCGTTTTTTT TTTCGTCTCG
GGGGGAGGCG GCGGCGGCCA TTGGTGGGGC GGGGAGAGGA GAAAGAAAGA GGCAAAAAA AAAGCAGAGC

981 GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCG CCCAGATCGG TGC GCGGGAG
CAGAGCTAGA AACCGGAACC ATCAAACCCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACGCGCCCTC

BamHI

~~~~~

1051 GGGCGGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCCGATCCT CGCGGGGAAT GGGGCTCTCG  
CCCGCCCTAG AGCGCCGACC GCAGAGGCC GCACTCAGCC GGGCCTAGGA GCGCCCCTTA CCCCAGAGAGC

BglII

~~~~~

1121 GATGTAGATC TTCTTTCTTT CTTCTTTTGG TGGTAGAATT TGAATCCCTC AGCATTGTTC ATCGGTAGTT
CTACATCTAG AAGAAAGAAA GAAGAAAAAC ACCATCTTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

HindIII NcoI

~~~~~

1191 TTTCTTTTCA TGATTTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTTTGT AG**GT**AGAAGC TTACCATGG  
AAAGAAAAGT ACTAAACACT GTTTACGTCG GAGCACGCCT CGAAAAAACA TC**CATC**TTTCG AATGGTACC

KpnI-EcoRI - deletion underlined and restored NCO site in bold in vectors pJQ4.9,  
pJQ3.2 and pJO6.3.

**FIG.\_50B**

153 / 154

**ALEURAIN\_d I t d NPIR (Ap plast) Structur and Sequ nc**



**ALEURAIN-NPIR-DEL**  
93 bp

+1            M   A   H   A   R   V   L   L   L   A   L   A   V   L   A   T   A   A   V   A  
HindIII NcoI

~~~~~  
1 AAGCTTACCA TGGCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG
TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC

+1 V A S S R A A
NotI

~~~~~  
71   TCGCCTCCTC CCGCGCGGCC GCC  
AGCGGAGGAG GGCGCGCCGG CGG

**FIG.\_51**

154 / 154

**SEE1 ( Senescence enhanced ) PROMOTER sequence**

```

1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI              XbaI      XbaI
      ~~~~~
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCggcacc GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATG

```

**FIG.\_52**

**SEE1 ( Senescence enhanced ) PROMOTER plus vacuolar aleurain SIGNAL/NPIR sequence**

```

1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI              XbaI      XbaI
      ~~~~~
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCggcacc GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
      M A H G R I L F L A L A V L
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
      BssHII
      ~~~~~
 NotI
      ~~~~~
      A T A A V A A A S L A D S N P I R P V T E R A
911 TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
      NotI
      ~~~~~
 A A
981 GGCCGCC

```

**FIG.\_53**